Effective Programs in Elementary Mathematics: A Best-Evidence Synthesis

Robert E. Slavin and Cynthia Lake

REVIEW OF EDUCATIONAL RESEARCH 2008; 78; 427
DOI: 10.3102/0034654308317473

Review of Educational Research
September 2008, Vol. 78, No. 3, pp. 427–515
DOI: 10.3102/0034654308317473

Effective Programs in Elementary Mathematics: A Best-Evidence Synthesis

Robert E. Slavin and Cynthia Lake
Johns Hopkins University

This article reviews research on the achievement outcomes of three types of approaches to improving elementary mathematics: mathematics curricula, computer-assisted instruction (CAI), and instructional process programs. Study inclusion requirements included use of a randomized or matched control group, a study duration of at least 12 weeks, and achievement measures not inherent to the experimental treatment. Eighty-seven studies met these criteria, of which 36 used random assignment to treatments. There was limited evidence supporting differential effects of various mathematics textbooks. Effects of CAI were moderate. The strongest positive effects were found for instructional process approaches such as forms of cooperative learning, classroom management and motivation programs, and supplemental tutoring programs. The review concludes that programs designed to change daily teaching practices appear to have more promise than those that deal primarily with curriculum or technology alone.

Keywords: best-evidence syntheses, elementary mathematics, experiments, meta-analysis, reviews of research.

The mathematics achievement of American children is improving but still has a long way to go. On the National Assessment of Educational Progress (NAEP, 2005), the math scores of fourth graders steadily improved from 1990 to 2005, increasing from 12% proficient or above to 35%. Among eighth graders, the percentage of students scoring proficient or better gained from 15% in 1990 to 20% in 2005. These trends are much in contrast to the trend in reading, which changed only slightly between 1992 and 2005.

However, although mathematics performance has grown substantially for all subgroups, the achievement gap between African American and Hispanic students and their White counterparts remains wide. In 2005, 47% of White fourth graders scored at or above proficient on the NAEP, but only 13% of African American students and 19% of Hispanic students scored this well.

Furthermore, the United States remains behind other developed nations in international comparisons of mathematics achievement. For example, U.S. 15-year-olds ranked 28th on the Organization for Economic Cooperation and Development's

Program for International Student Assessment study of mathematics achievement, significantly behind countries such as Finland, Canada, the Czech Republic, and France.

Although we can celebrate the growth of America's children in mathematics, we cannot be complacent. The achievement gap between children of different ethnicities, and between U.S. children and those in other countries, gives us no justification for relaxing our focus on improving mathematics for all children. Under No Child Left Behind, schools can meet their adequate yearly progress goals only if all subgroups meet state standards (or show adequate growth) in all subjects tested. Nationally, thousands of schools are under increasing sanctions because the school, or one or more subgroups, is not making sufficient progress in math. For this reason, educators are particularly interested in implementing programs and practices that have been shown to improve the achievement of all children.

One way to reduce mathematics achievement gaps, and to improve achievement overall, is to provide low-performing schools training and materials known to be markedly more effective than typical programs. No Child Left Behind, for example, emphasizes the use of research-proven programs to help schools meet their adequate yearly progress goals. Yet, for such a strategy to be effective, it is essential to know what specific programs are most likely to improve mathematics achievement. The purpose of this review is to summarize and interpret the evidence on elementary mathematics programs in hopes of informing policies and practices designed to reduce achievement gaps and improve the mathematics achievement of all children.

Reviews of Effective Programs in Elementary Mathematics

The No Child Left Behind Act strongly emphasizes encouraging schools to use federal funding (such as Title I) on programs with strong evidence of effectiveness from "scientifically-based research." No Child Left Behind defines scientifically based research as research that uses experimental methods to compare groups using programs to control groups, preferably with random assignment to conditions. Yet, in mathematics, what programs meet this standard? There has never been a published review of scientific research on all types of effective programs. There have been meta-analyses on outcomes of particular approaches to mathematics education, such as use of educational technology (e.g., H. J. Becker, 1992; Chambers, 2003; Kulik, 2003; R. Murphy et al., 2002), calculators (e.g., Ellington, 2003), and math approaches for at-risk children (e.g., Baker, Gersten, & Lee, 2002; Kroesbergen & Van Luit, 2003). There have been reviews of the degree to which various math programs correspond to current conceptions of curriculum, such as those carried out by Project 2061 evaluating middle school math textbooks (American Association for the Advancement of Science, 2000). The What Works Clearinghouse (2006) is doing a review of research on effects of alternative math textbooks, but this has not appeared as of this writing. However, there are no comprehensive reviews of research on all of the programs and practices available to educators.

In 2002, the National Research Council (NRC) convened a blue-ribbon panel to review evaluation data on the effectiveness of mathematics curriculum materials, focusing in particular on innovative programs supported by the National Science Foundation (NSF) but also looking at evaluations of non-NSF materials (Confrey, 2006; NRC, 2004). The NRC panel assembled research evaluating elementary and secondary math programs and ultimately agreed on 63 quasi-experimental studies

covering all grade levels, kindergarten through 12, that they considered to meet minimum standards of quality.

The authors of the NRC (2004) report carefully considered the evidence across the 63 studies and decided that they did not warrant any firm conclusions. Using a vote-count procedure, they reported that among 46 studies of innovative programs that had been supported by NSF, 59% found significantly positive effects, 6% significant negative effects, and 35% no differences. Most of these studies involved elementary and secondary programs of the University of Chicago School Mathematics Project. For commercial programs, the corresponding percentages were 29%, 13%, and 59%. Based on this, the report tentatively suggested that NSF-funded programs had better outcomes. Other than this very general finding, the report was silent about the evidence on particular programs. In addition to concerns about methodological limitations, the report maintained that it is not enough to show differences in student outcomes; curricula, the authors argued, should be reviewed for content by math educators and mathematicians to be sure they correspond to current conceptions of what math content should be. None of the studies combined this kind of curriculum review with rigorous evaluation methods, so the NRC chose not to describe the outcomes it found in the 63 evaluations that met its minimum standards (see Confrey, 2006).

Focus of the Present Review

This review examines research on all types of math programs that are available to elementary educators today. The intention is to place all types of programs on a common scale. In this way, we hope to provide educators with meaningful, unbiased information that they can use to select programs and practices most likely to make a difference with their students. In addition, the review is intended to look broadly for factors that might underlie effective practices across programs and program types and to inform an overarching theory of effective instruction in elementary mathematics.

The review examines three general categories of math approaches. One is *mathematics curricula,* in which the main focus of the reform is on introduction of alternative textbooks. Some of the programs in this category were developed with extensive funding by the NSF, which began in the 1990s. Such programs provide professional development to teachers, and many include innovative instructional methods. However, the primary theory of action behind this set of reforms is that higher level objectives, including a focus on developing critical mathematics concepts and problem-solving skills and pedagogical aids such as the use of manipulatives and improved sequencing of objectives, and other features of textbooks will improve student outcomes. Outcomes of such programs have not been comprehensively reviewed previously, although they are currently being reviewed by the What Works Clearinghouse (2006).

A second major category of programs is *computer-assisted instruction* (CAI), which uses technology to enhance student achievement in mathematics. CAI programs are almost always supplementary, meaning that students experience a full textbook-based program in mathematics and then go to a computer lab or a classroom-based computer to receive additional instruction. CAI programs diagnose students' levels of performance and then provide exercises tailored to students' individual needs. Their theory of action depends substantially on this individualization and on the

computer's ability to continuously assess students' progress and accommodate their needs. This is the one category of program that has been extensively reviewed in the past, most recently by Kulik (2003), R. Murphy et al. (2002), and Chambers (2003).

A third category is *instructional process* programs. This set of interventions is highly diverse, but what characterizes its approaches is a focus on teachers' instructional practices and classroom management strategies rather than on curriculum or technology. With two exceptions, studies of instructional process strategies hold curriculum constant. Instructional process programs introduce variations in within-class grouping arrangements (as in cooperative learning or tutoring) and in the amounts and uses of instructional time. Their theories of action emphasize enhancing teachers' abilities to motivate children, to engage their thinking processes, to improve classroom management, and to accommodate instruction to students' needs. Their hallmark is extensive professional development, usually incorporating follow-up and continuing interactions among the teachers themselves.

The three approaches to mathematics reform can be summarized as follows:

- change the curriculum,
- supplement the curriculum with CAI, or
- change classroom practices.

The categorization of programs in this review relates to a long-standing debate in research on technology by Kozma (1994) and Clark (2001). Clark argued that research on technology must hold curriculum constant in order to identify the unique contributions of the technology. Kozma replied that technology and curriculum were so intertwined that it was not meaningful to separate them in analysis. As a practical matter, content, media, and instructional processes are treated in different ways in the research discussed here. The mathematics curricula vary textbooks but otherwise do not make important changes in media or instructional methods. The CAI studies invariably consider technology and curricula together; none do as Clark suggested. Most of the instructional process studies vary only the teaching methods and professional development, holding curriculum constant, but a few (Team-Assisted Individualization [TAI] Math, Project CHILD, Direct Instruction, and Project SEED) combine curricula, processes, and (in the case of Project CHILD) media.

The categorization of the programs was intended to facilitate understanding and contribute to theory, not to restrict the review. No studies were excluded due to lack of fit with one of the categories. This review examines research on dozens of individual programs to shed light on the broad types of mathematics reforms most likely to enhance the mathematics achievement of elementary school children.

Review Method

This article reviews studies of elementary mathematics programs in an attempt to apply consistent, well-justified standards of evidence to draw conclusions about effective elementary mathematics programs. The review applies a technique called "best-evidence synthesis" (Slavin, 1986, 2007), which seeks to apply consistent, well-justified standards to identify unbiased, meaningful quantitative information from experimental studies, and then discusses each qualifying study, computing

effect sizes but also describing the context, design, and findings of each. Best-evidence synthesis closely resembles meta-analysis (Cooper, 1998; Lipsey & Wilson, 2001), but it requires more extensive discussion of key studies instead of primarily pooling results across many studies. In reviewing educational programs, this distinction is particularly important, as there are typically few studies of any particular program, so understanding the nature and quality of the contribution made by each study is essential. The review procedures, described below, are similar to those applied by the What Works Clearinghouse (2005, 2006). (See Slavin, 2007, for a detailed description and justification for the review procedures used here and in syntheses by Cheung & Slavin, 2005, and by Slavin, Lake, & Groff, 2007.)

The purpose of this review is to examine the quantitative evidence on elementary mathematics programs to discover how much of a scientific basis there is for competing claims about the effects of various programs. Our intention is to inform practitioners, policy makers, and researchers about the current state of the evidence on this topic and to identify gaps in the knowledge base that are in need of further scientific investigation.

Limitations of the Review

This article is a quantitative synthesis of achievement outcomes of alternative mathematics approaches. It does not report on qualitative or descriptive evidence, attitudes, or other nonachievement outcomes. These are excluded not because they are unimportant but because space limitations do not allow for a full treatment of all of the information available on each program. Each report cited, and many that were not included (listed in Appendix A), contain much valuable information, such as descriptions of settings, nonquantitative and nonachievement outcomes, and the story of what happened in each study. The present article extracts from these rich sources just the information on experimental control differences on quantitative achievement measures in order to contribute to an understanding of the likely achievement affects of using each of the programs discussed. Studies are included or excluded and are referred to as being high or low in quality solely based on their contributions to an unbiased, well-justified quantitative estimate of the strength of the evidence supporting each program. For a deeper understanding of all of the findings of each study, please see the original reports.

Literature Search Procedures

A broad literature search was carried out in an attempt to locate every study that could possibly meet the inclusion requirements. This included obtaining all of the elementary studies cited by the NRC (2004) and by other reviews of mathematics programs, including technology programs that teach math (e.g., Chambers, 2003; Kulik, 2003; R. Murphy et al., 2002). Electronic searches were made of educational databases (JSTOR, ERIC, EBSCO, PsycINFO, and Dissertation Abstracts), Web-based repositories (Google, Yahoo, Google Scholar), and math education publishers' Web sites. Citations of studies appearing in the studies found in the first wave were also followed up.

Effect Sizes

In general, effect sizes were computed as the difference between experimental and control individual student posttests, after adjustment for pretests and other

covariates, divided by the unadjusted control group standard deviation. If the control group standard deviation was not available, a pooled standard deviation was used. Procedures described by Lipsey and Wilson (2001) and by Sedlmeier and Gigerenzer (1989) were used to estimate effect sizes when unadjusted standard deviations were not available, as when the only standard deviation presented was already adjusted for covariates or when only gain score standard deviations were available. School- or classroom-level standard deviations were adjusted to approximate individual-level standard deviations, as aggregated standard deviations tend to be much smaller than individual standard deviations. If pretest and posttest means and standard deviations were presented but adjusted means were not, effect sizes for pretests were subtracted from effect sizes for posttests.

Criteria for Inclusion

Criteria for inclusion of studies in this review were as follows:

1. The studies involved elementary (K–5) children, plus sixth graders if they were in elementary schools.
2. The studies compared children taught in classes using a given mathematics program to those in control classes using an alternative program or standard methods.
3. Studies could have taken place in any country, but the report had to be available in English.
4. Random assignment or matching with appropriate adjustments for any pretest differences (e.g., ANCOVA) had to be used. Studies without control groups, such as pre–post comparisons and comparisons to "expected" gains, were excluded. Studies with pretest differences of more than 50% of a standard deviation were excluded, because even with ANCOVAs, large pretest differences cannot be adequately controlled for, as underlying distributions may be fundamentally different. (See the Methodological Issues section, later in the article, for a discussion of randomized and matched designs.)
5. The dependent measures included quantitative measures of mathematics performance, such as standardized mathematics measures. Experimenter-made measures were accepted if they were described as comprehensive measures of mathematics, which would be fair to the control groups, but measures of math objectives inherent to the program (but unlikely to be emphasized in control groups) were excluded. For example, a study of CAI by Van Dusen and Worthen (1994) found no differences on a standardized test (effect size [ES] = +0.01) but a substantial difference on a test made by the software developer (ES = +0.35). The software-specific measure was excluded, as it probably focused on objectives and formats practiced in the CAI group but not in the control group. (See the Methodological Issues section, later in the article, for a discussion of this issue.)
6. A minimum treatment duration of 12 weeks was required. This requirement is intended to focus the review on practical programs intended for use for the whole year rather than on brief investigations. Brief studies may not allow programs intended to be used over the whole year to show their full effect. On the other hand, brief studies often advantage experimental groups that

focus on a particular set of objectives during a limited time period while control groups spread that topic over a longer period. For example, a 30-day experiment by Cramer, Post, and delMas (2002) evaluated a fractions curriculum that is part of the Rational Number Project. Control teachers using standard basals were asked to delay their fractions instruction to January to match the exclusive focus of the experimental group on fractions, but it seems unlikely that their focus would have been equally focused on fractions, the only skill assessed.

Appendix A lists studies that were considered but excluded according to these criteria, as well as the reasons for exclusion. Appendix B lists abbreviations used throughout this review.

Methodological Issues in Studies of Elementary Mathematics Programs

The three types of mathematics programs reviewed here, mathematics curricula, CAI programs, and instructional process programs, suffer from different characteristic methodological problems (see Slavin, 2007). Across most of the evaluations, lack of random assignment is a serious problem. Matched designs are used in most studies that met the inclusion criteria, and matching leaves studies open to selection bias. That is, schools or teachers usually choose to implement a given experimental program and are compared to schools or teachers who did not choose the program. The fact of this self-selection means that no matter how well experimental groups and control groups are matched on other factors, the experimental group is likely to be more receptive to innovation, more concerned about math, have greater resources for reform, or otherwise have advantages that cannot be controlled for statistically. Alternatively, it is possible that schools that would choose a given program might have been dissatisfied with results in the past and might therefore be less effective than comparable schools. Either way, matching reduces internal validity by allowing for the possibility that outcomes are influenced by whatever (unmeasured) factors that led the school or teacher to choose the program. It affects external validity in limiting the generalization of findings to schools or teachers who similarly chose to use the program.

Garden-variety selection bias is bad enough in experimental design, but many of the studies suffered from design features that add to concerns about selection bias. In particular, many of the curriculum evaluations used a post hoc design, in which a group of schools using a given program, perhaps for many years, is compared after the fact to schools that matched the experimental program at pretest or that matched on other variables, such as poverty or reading measures. The problem is that only the "survivors" are included in the study. Schools that bought the materials, received the training, but abandoned the program before the study took place are not in the final sample, which is therefore limited to more capable schools. As one example of this, Waite (2000), in an evaluation of Everyday Mathematics, described how 17 schools in a Texas city originally received materials and training. Only 7 were still implementing it at the end of the year, and 6 of these agreed to be in the evaluation. We are not told why the other schools dropped out, but it is possible that the staff members of the remaining 6 schools may have been more capable or motivated than those at schools that dropped the program.

The comparison group in the same city was likely composed of the full range of more and less capable school staff members, and they presumably had the same opportunity to implement Everyday Mathematics but chose not to do so. Other post hoc studies, especially those with multiyear implementations, must also have had some number of dropouts but typically did not report how many schools there were at first and how many dropped out. There are many reasons schools may have dropped out, but it seems likely that any school staff able to implement any innovative program for several years is more capable, more reform oriented, or better led than those unable to do so or (even worse) than those that abandoned the program because it was not working. As an analog, imagine an evaluation of a diet regimen that studied only people who kept up the diet for a year. There are many reasons a person might abandon a diet, but chief among them is that it is not working, so looking only at the nondropouts would bias such a study.

Worst of all, post hoc studies usually report outcome data selected from many potential experimental and comparison groups and may therefore report on especially successful schools using the program or matched schools that happen to have made particularly small gains, making an experimental group look better by comparison. The fact that researchers in post hoc studies often have pretest and posttest data readily available on hundreds of potential matches, and may deliberately or inadvertently select the schools that show the program to best effect, means that readers must take results from after-the-fact comparisons with a grain of salt.

Finally, because post hoc studies can be very easy and inexpensive to do, and are usually contracted for by publishers rather than supported by research grants or conducted for dissertations, such studies are likely to be particularly subject to the "file drawer" problem. That is, post hoc studies that fail to find expected positive effects are likely to be quietly abandoned, whereas studies supported by grants or produced for dissertations will almost always result in a report of some kind. The file drawer problem has been extensively described in research on meta-analyses and other quantitative syntheses (see, e.g., Cooper, 1998), and it is a problem in all research reviews, but it is much more of a problem with post hoc studies.

Despite all of these concerns, post hoc studies were reluctantly included in this review for one reason: Without them, there would be no evidence at all concerning most of the commercial textbook series used by the vast majority of elementary schools. As long as the experimental and control groups were well matched at pretest on achievement and demographic variables, and met other inclusion requirements, we decided to include them, but readers should be very cautious in interpreting their findings. Prospective studies, in which experimental and control groups were designated in advance and outcomes are likely to be reported whatever they turn out to be, are always to be preferred to post hoc studies, other factors being equal.

Another methodological limitation of almost all of the studies in this review is analysis of data at the individual-student level. The treatments are invariably implemented at the school or classroom levels, and student scores within schools and classrooms cannot be considered independent. In clustered settings, individual-level analysis does not introduce bias, but it does greatly overstate statistical significance, and in studies involving a small number of schools or classes it can cause treatment effects to be confounded with school or classroom effects. In an extreme form, a study comparing, say, one school or teacher using Program A and one using

Program B may have plenty of statistical power at the student level, but treatment effects cannot be separated from characteristics of the schools or teachers.

Several studies did randomly assign groups of students to treatments but nevertheless analyzed at the individual-student level. The random assignment in such studies is beneficial because it essentially eliminates selection bias. However, analysis at the student level, rather than at the level of random assignment, still confounds treatment effects and school or classroom effects, as noted earlier. We call such studies *randomized quasi-experiments* and consider them more methodologically rigorous, all other things being equal, than matched studies, but less so than randomized studies in which analysis is at the level of random assignment.

Some of the qualifying studies, especially of instructional process programs, were quite small, involving a handful of schools or classes. Beyond the problem of confounding, small studies often allow the developers or experimenters to be closely involved in implementation, creating far more faithful and high-quality implementations than would be likely in more realistic circumstances. Unfortunately, many of the studies that used random assignment to treatments were very small, often with just one teacher or class per treatment. Also, the file drawer problem is heightened with small studies, which are likely to be published or otherwise reported only if their results are positive (see Cooper, 1998).

Another methodological problem inherent to research on alternative mathematics curricula relates to outcome measures. In a recent criticism of the What Works Clearinghouse, Schoenfeld (2006) expressed concern that because most studies of mathematics curricula use standardized tests or state accountability tests focused more on traditional skills than on concepts and problem solving, there is a serious risk of "false negative" errors, which is to say that studies might miss true and meaningful effects on unmeasured outcomes characteristic of innovative curricula (also see Confrey, 2006, for more on this point). This is indeed a serious problem, and there is no solution to it. Measuring content taught only in the experimental group risks false positive errors, just as use of standardized tests risks false negatives. In the present review, only outcome measures that assess content likely to have been covered by all groups are considered; measures inherent to the treatment are excluded. However, many curriculum studies include outcomes for subscales, such as computation, concepts and applications, and problem solving, and these outcomes are separately reported in this review. Therefore, if an innovative curriculum produces, for example, better outcomes on problem solving but no differences on computation, that might be taken as an indication that it is succeeding at least in its area of emphasis.

A total of 87 studies met the inclusion criteria. Tables 1, 2, and 3 list all the qualifying studies. Within sections on each program, studies that used random assignment (if any) are listed first, then randomized quasi-experiments, then prospective matched studies, and finally post hoc matched studies. Within these categories, studies with larger sample sizes are listed first.

This article discusses conclusions drawn from the qualifying studies, but study-by-study discussions are withheld so that the article will fit within the space requirements of the *Review of Educational Research.* These descriptions appear in a longer version of this article, which can be found at http://www.bestevidence.org/_images/word_docs/Eff%20progs%20ES%20math%20Version%201.2%20for%20BEE%2002%2009%2007.doc.

Studies of Mathematics Curricula

Perhaps the most common approach to reform in mathematics involves adoption of reform-oriented textbooks, along with appropriate professional development. Programs that have been evaluated fall into three categories. One is programs developed under funding from the NSF that emphasize a constructivist philosophy, with a strong emphasis on problem solving, manipulatives, and concept development and a relative de-emphasis on algorithms. At the opposite extreme is Saxon Math, a back-to-basics curriculum that emphasizes building students' confidence and skill in computation and word problems. Finally, there are traditional commercial textbook programs.

The reform-oriented programs supported by NSF, especially Everyday Mathematics, have been remarkably successful in making the transition to widespread commercial application. Sconiers, Isaacs, Higgins, McBride, and Kelso (2003) estimated that, in 1999, 10% of all schools were using one of three programs that had been developed under NSF funding and then commercially published. That number is surely higher as of this writing. Yet, experimental control evaluations of these and other curricula that meet the most minimal standards of methodological quality are very few. Only five studies of the NSF programs met the inclusion standards, and all but one of these was a post hoc matched comparison.

This section reviews the evidence on mathematics curricula. Overall, 13 studies met the inclusion criteria, of which only 2 used random assignment. Table 1 summarizes the methodological characteristics and outcomes of these studies. Descriptions of each study can be seen at http://www.bestevidence.org/_images/word_docs/Eff%20progs%20ES%20math%20Version%201.2%20for%20BEE%2002%2009%2007.doc.

With a few exceptions, the studies that compared alternative mathematics curricula are of marginal methodological quality. Ten of the 13 qualifying studies used post hoc matched designs in which control schools, classes, or students were matched with experimental groups after outcomes were known. Even though such studies are likely to overstate program outcomes, the outcomes reported in these studies are modest. The median effect size was only +0.10. The enormous ARC study (Sconiers et al., 2003) found an average effect size of only +0.10 for the three most widely used of the NSF-supported mathematics curricula, taken together. Riordan and Noyce (2001), in a post hoc study of Everyday Mathematics, did find substantial positive effects (ES = +0.34) in comparison to controls for schools that had used the program for 4 to 6 years, but effects for schools that used the program for 2 to 3 years were much smaller (ES = +0.15). This finding may suggest that schools need to implement this program for 4 to 6 years to see a meaningful benefit, but the difference in outcomes may just be a selection artifact, due to the fact that schools that were not succeeding may have dropped the program before the 4th year. The evidence for the impacts of all of the curricula on standardized tests is thin. The median effect size across five studies of the NSF-supported curricula is only +0.12, very similar to the findings of the ARC study.

The reform-oriented math curricula may have positive effects on outcomes not assessed by standardized tests, as suggested by Schoenfeld (2006) and Confrey (2006). However, the results on standardized and state accountability measures do not suggest differentially strong impacts on outcomes such as problem solving or concepts and applications that one might expect, as these are the focus of the NSF curricula and other reform curricula.

(text continues on p. 445)

TABLE 1
Mathematics curricula: Descriptive information and effect sizes for qualifying studies

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Everyday Mathematics; Math Trailblazers; and Investigations in Number, Data, and Space (ARC Study)									
Sconiers, Isaacs, Higgins, McBride, & Kelso (2003	Matched post hoc	1 year	742 schools, 100,875 students	3-5	Schools across Illinois, Massachusetts, and Washington	Matched on reading scores, SES, and ethnicity	ISAT/MCAS/ ITBS/WASL		+0.10
							Computation	+0.10	
							Measurement	+0.14	
							Geometry	+0.08	
							Probability/ statistics	+0.03	
							Algebra	+0.09	
							Race/ethnicity		
							Asian	+0.11	
							Black	+0.09	
							Hispanic	+0.02	
							White	+0.10	
							SES		
							Low	+0.11	
							Middle	+0.08	
							High	+0.10	

(continued)

TABLE 1 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Everyday Mathematics									
Woodward & Baxter (1997)	Matched	1 year	3 schools, 38 students	3	Middle-class, suburban schools in the Pacific Northwest. Students scoring below 34th percentile.	Matched on pretests	ITBS		–0.25
							Computation	–0.22	
							Concepts	+0.10	
							Problem solving	–0.10	
SRA/McGraw Hill (2003)	Matched post hoc	1 year	562 schools, 39,701 students	3-5	Schools across Illinois, Massachusetts, and Washington (subset of Sconiers et al., 2003, study).	Matched on reading scores, SES, and ethnicity	ISAT/MCAS/ ITBS/WASL		+0.12
							Computation	+0.13	
							Measurement	+0.15	
							Geometry	+0.12	
							Probability	+0.04	
							Algebra	+0.07	
							Problem solving	+0.05	
							Race/ ethnicity		
							Asian	+0.11	
							Black	+0.11	
							Hispanic	+0.02	
							White	+0.13	

(continued)

TABLE 1 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
							SES		
							Low	+0.14	
							Middle	+0.09	
							High	+0.10	
Riordan & Noyce (2001)	Matched post hoc	2-4+ years	145 schools, 8,793 students	4	Schools across Massachusetts. Mostly White, non-free lunch.	Matched with controls on prior state tests, SES.	MCAS		+0.25
							2-3 years of Everyday Math	+0.15	
							4+ years of Everyday Math	+0.34	
Waite (2000)	Matched post hoc	1 year	Schools: 6 Everyday Math, 12 control; students: 732 Everyday Math, 2,704 control	3-5	Urban district in northern Texas	Matched on prior mathematics test, SES, and ethnicity.	TAAS		+0.26
							Operations	+0.25	
							Problem solving	+0.31	
							Concepts	+0.24	
							Race/ethnicity		
							White	+0.33	
							Black	+0.34	
							Hispanic	0.00	
							SES		
							Low	+0.27	
							Middle high	+0.18	
							Grade		
							Grade 3	+0.18	
							Grade 4	+0.12	
							Grade 5	+0.29	

(continued)

TABLE 1 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Math Trailblazers									
Kelso, Booton, & Majumdar (2003)	Matched post hoc	1 year	4,942 students	3, 4	Schools across Washington State (subset of Sconiers et al., 2003, study)	Matched with controls on reading, SES, and ethnicity.	ITBS/WASL Computation Concepts Problem solving	 0.00 +0.09 +0.07	+0.06
Saxon Math									
Resendez & Azin (2005	Matched post hoc	1-5 years	340 schools	1-5	Georgia public schools	Matched based on SES, race/ethnicity	Criterion-Referenced Competency Test. (Georgia)		+0.02

(continued)

TABLE 1 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Scott Foresman-Addison Wesley									
Resendez & Sridharan (2005, 2006) and Resendez & Manley (2005)	Random assignment	1 year	4 schools, 39 teachers, 901 students	3, 5	Schools in Ohio and New Jersey	Random assignment within schools. Demographic variables and prior achievement used as covariates.	Terra Nova Math Total Computation	 –0.07 +0.05	–0.01
Resendez & Sridharan (2005) and Resendez & Azin (2005)	Random assignment	1 year	6 schools, 35 teachers, 719 students	2, 4	Schools across Wyoming, Washington, Kentucky, and Virginia	Random assignment was done within schools. Pretests used as covariates.	Terra Nova Math Total Computation	+0.11 –0.04 +0.05	+0.04

(continued)

TABLE 1 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Houghton Mifflin									
J. Johnson, Yanyo, & Hall (2002)	Matched post hoc	1 year	16 districts, 297 schools	2-5	Districts across California	Matched on prior math achievement, demographic variables, and district sizes.	SAT-9		+0.14
Growing With Mathematics									
Biscoe & Harris (2004)	Matched post hoc	1 year	144 classrooms	K-5	Schools across Arkansas, Hawaii, Iowa, Oklahoma, and New Jersey	Matched on grade, race, and math pretests.	Terra Nova Comprehension Computation	 +0.20 +0.23	+0.22

(continued)

TABLE 1 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup	Overall effect size
Excel Math								
Mahoney (1990)	Matched post hoc	1 year	6 schools; students: 221 experimental, 273 control	2, 4	Schools in Palm Springs, California	Matched on SES, ethnicity, and achievement.	SAT	+0.13
Math Steps								
Chase, Johnston, Delameter, Moore, & Golding (2000)	Matched post hoc	1 year	Students: 2,422 treatment, 1,805 control	3-5	Schools across five California school districts	Matched on school characteristics, demographics, and math achievement.	SAT-9	+0.03

(continued)

TABLE 1 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
							Grade 3	+0.11	
							Grade 4	+0.03	
							Grade 5	−0.04	
Knowing Mathematics									
Houghton Mifflin (n.d.)	Matched post hoc	12 weeks	4 schools, 39 students	4-6	Schools in Lincoln, Nebraska	Matched on demographics and prior math achievement.	MAT-8		+0.10
							Computation	+0.20	
							Problem solving	+0.14	

Note. See Appendix B for the table of abbreviations.

Evidence supporting Saxon Math, the very traditional, algorithmically focused curriculum that is the polar opposite of the NSF-supported models, was lacking. The one methodologically adequate study evaluating the program, by Resendez and Azin (2005), found no differences on Georgia state tests between elementary students who experienced Saxon Math and those who used other texts.

A review of research on middle and high school math programs by Slavin et al. (2007) using methods essentially identical to those used in the present review also found very limited impacts of alternative textbooks. Across 38 qualifying studies, the median effect size was only +0.05. The effect sizes were also +0.05 in 24 studies of NSF-supported curricula and were +0.12 in 11 studies of Saxon Math.

More research is needed on all of these programs, but the evidence to date suggests a surprising conclusion that despite all the heated debates about the content of mathematics, there is limited high-quality evidence supporting differential effects of different math curricula.

Computer-Assisted Instruction

A long-standing approach to improving the mathematics performance of elementary students is computer-assisted instruction, or CAI. Over the years, CAI strategies have evolved from limited drill-and-practice programs to sophisticated integrated learning systems that combine computerized placement and instruction. Typically, CAI materials have been used as supplements to classroom instruction and are often used only a few times a week. Some of the studies of CAI in math have involved only 30 minutes per week. What CAI primarily adds is the ability to identify children's strengths and weaknesses and then give them self-instructional exercises designed to fill in gaps. In a hierarchical subject like mathematics, especially computation, this may be of particular importance.

A closely related strategy, computer-managed learning systems, is also reviewed in this section as a separate subcategory.

As noted earlier, CAI is one of the few categories of elementary mathematics interventions that has been reviewed extensively. Most recently, for example, Kulik (2003) reviewed research on the uses of CAI in reading and math and concluded that studies supported the effectiveness of CAI for math but not for reading. R. Murphy et al. (2002) concluded that CAI was effective in both subjects, but with much larger effects in math than in reading. A recent large, randomized evaluation of various CAI programs found no effects on student achievement in reading or math (Dynarski et al., 2007).

Table 2 summarizes qualifying research on several approaches to CAI in elementary mathematics. Many of these involved earlier versions of CAI that no longer exist, but it is still useful to be aware of the earlier evidence, as many of the highest quality studies were done in the 1980s and early 1990s. Overall, 38 studies of CAI met the inclusion criteria, and 15 of these used randomized or randomized quasi-experimental designs. In all cases, control groups used nontechnology approaches, such as traditional textbooks. For descriptions of each study, see http://www.bestevidence.org/_images/word_docs/Eff%20progs%20ES%20math%20Version%201.2%20for%20BEE%2002%2009%2007.doc.

In sheer numbers of studies, CAI is the most extensively studied of all approaches to elementary math reform. Most studies of CAI find positive effects, especially on

(text continues on p. 459)

TABLE 2
Computer-assisted instruction: Descriptive information and effect sizes for qualifying studies

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Jostens/Compass Learning									
Alifrangis (1991)	Random assignment	1 year	1 school, 12 classes, 250 students	4-6	School at an army base near Washington, DC	Random assignment, matched on pretests.	CTBS		–0.08
H. J. Becker (1994)	Random assignment	1 year	1 school, 8 classes	2-5	Inner-city East Coast school	Random assignment, matched on pretests.	CAT		+0.07
							Computation	+0.10	
							Concepts and applications	–0.02	
Zollman, Oldham, & Wyrick (1989)	Matched	1 year	15 schools; students: 146 treatment, 274 control	4-6	Lexington, Kentucky. Chapter 1 students.	Matched on Chapter 1 status.	MAT-6		+0.21

(continued)

TABLE 2 (continued)

Study	Design	Duration	N	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Hunter (1994)	Matched	1 year	6 schools, 150 students	2-5	Public schools in Jefferson County, Georgia. Chapter 1 students.	Matched on ethnicity and SES.	ITBS		+0.40
Estep, McInerney, Vockell, & Kosmoski (1999-2000)	Matched post hoc	1-5 years	106 schools	3	Schools across Indiana	Matched on ISTEP pretests.	ISTEP		+0.01
Spencer (1999)	Matched post hoc	5 years	92 students	2, 3	Urban school district in southeastern Michigan	Matched on gender, race, and past CAT total math scores.	CAT		+0.40
							Grade 2 Starters	+0.37	
							Grade 3 Starters	+0.44	

(continued)

TABLE 2 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Clariana (1994)	Matched post hoc	1 year	1 school, 4 classes, 85 students	3	School in a predominantly White, rural area	Matched on pretests.	CTBS		+0.66
CCC/Successmaker									
Ragosta (1983)	Random assignment	3 years	4 schools	1-6	Schools in the Los Angeles Unified School District	Random assignment, controlled for pretests, sex, ethnicity.	CTBS Computations Concepts Applications	 +0.72 +0.09 +0.26	
Hotard & Cortez (1983)	Random assignment	6 months	2 schools, 190 students	3-6	Schools in Lafayette Parish, Louisiana	Random assignment, matched on pretests.	CTBS		+0.19
Manuel (1987)	Random assignment	12 weeks	3 schools, 165 students	3-6	Schools in Omaha, Nebraska	Random assignment, matched on pretests.	CTBS		+0.07

(continued)

TABLE 2 (continued)

Study	Design	Duration	N	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup	Overall effect size
Mintz (2000)	Matched post hoc	1 year	8 schools, 487 students	4, 5	Schools in Etowah County, Alabama	Matched on School Ability Index	SAT-9	–0.06
Laub (1995)	Matched post hoc	5 months	2 schools, 14 classes, 314 students	4, 5	Schools in Lancaster County, Pennsylvania	Matched on SAT Total Math.	SAT	+0.27
Classworks								
Patterson (2005)	Random assignment	14 weeks	30 students	3	Rural school in central Texas	Matched on pretests.	SAT-9	+0.85
Whitaker (2005)	Matched post hoc	1 year	2 schools, 218 students	4, 5	Schools in rural Tennessee	Matched on demographics and pretests.	TCAP	+0.21

(continued)

TABLE 2 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Lightspan									
Birch (2002)	Matched post hoc	2 years	2 schools, 101 students	2, 3	Schools in the Caesar Rodney School District in Delaware	Matched on pretests, demographics.	SAT		+0.28
							End of Year 1	+0.53	
							End of Year 2	+0.28	
Other CAI									
H. J. Becker (1994) (computer networking specialists)	Random assignment	1 year	1 school, 9 classes	2-5	Inner-city East Coast school	Random assignment, matched on pretests.	CAT		+0.18
							Computation	+0.18	
							Concepts and applications	+0.12	

(continued)

TABLE 2 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Carrier, Post, & Heck (1985) (various CAI)	Random assignment	14 weeks	6 classes, 144 students	4	Metropolitan school district in Minnesota	Random assignment, matched on pretests.	Experimenter-designed test		+0.21
							Symbolic algorithms	+0.01	
							Division facts	+0.28	
							Multiplication facts	+0.35	
Abram (1984) (The Math Machine)	Random assignment	12 weeks	1 school, 5 classes, 103 students	1	Suburban school district in Southwest	Random assignment	ITBS		–0.18
Watkins (1986) (The Math Machine)	Random assignment	6 months	1 school, 82 students	1	Suburban Southwestern school	ITBS and Cognitive Abilities Test served as covariates.	CAT		+0.41

(continued)

TABLE 2 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Fletcher, Hawley, & Piele (1990) (Milliken Math Sequences)	Random assignment	4 months	1 school, 4 classes, 79 students	3, 5	School in rural Saskatchewan	Matched on pretests.	Canadian Test of Basic Skills		+0.40
							Grade 3	+0.48	
							Computation	+0.58	
							Concepts	+0.20	
							Problem solving	+0.10	
							Grade 5	+0.32	
							Computation	+0.22	
							Concepts	+0.30	
							Problem solving	+0.36	
Van Dusen & Worthen (1994) (unspecified program)	Randomized quasi-experiment	1 year	6 schools, 141 classes, 4,612 students	K-6	Schools selected from diverse geographic areas across the United States	Random assignment, controlled for pretests.	Norm-referenced tests		+0.01
							Good implementers	+0.05	
							Weak implementers	–0.04	

(continued)

TABLE 2 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup	Overall effect size
Shanoski (1986) (Mathematics Courseware)	Randomized quasi-experiment	20 weeks	4 schools, 32 classes, 832 students	2-6	Rural Pennsylvania	Matched on pretests.	CAT	–0.02
Turner (1985) (Milliken Math Sequencing and Pet Professor)	Randomized quasi-experiment	15 weeks	275 students	3-4	School in sub-urb of Phoenix, Arizona	Random assignment, matched on pretests.	CTBS	+0.37
Schmidt (1991) (Wasatch integrated learning system)	Matched	1 year	4 schools, 1,224 students	2-6	Schools in Southern California	Matched on SES and CTBS scores.	CTBS	+0.05

(continued)

TABLE 2 (continued)

Study	Design	Duration	N	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup	Overall effect size
Bass, Ries, & Sharpe (1986) (CICERO software)	Matched	1 year	1 school, 178 students	5-6	School in rural Virginia. Chapter 1 students.	Matched on pretests.	SRA Achievement Series	+0.02
A. H. Webster (1990) (Courses by Computers Math)	Matched	14 weeks	5 schools, 120 students	5	Schools in rural Mississippi Delta school district	Matched on pretests, demographics.	SAT	+0.13
Hess & McGarvey (1987) (Memory, Number Farm)	Matched	5 months	66 students	K	Schools drew students from a wide range of socioeconomic backgrounds.	Matched on pretests, gender.	Criterion-referenced test	+0.14

(continued)

TABLE 2 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup	Overall effect size
Gilman & Brantley (1988)	Matched	1 year	1 school, 57 students	4	School in rural Indiana	Matched on pretests.	ITBS	+0.03
Miller (1997) (Waterford Integrated Learning System)	Matched post hoc	1 to 3 years	Schools: 10 Waterford Integrated Learning System, 20 control	3-5	New York City Public Schools	Matched on pretests, demographics, and attendance.	MAT	+0.17
Levy (1985) (Mathematics Strands, Problem Solving—Instructional Systems, Incorporated)	Matched post hoc	1 year	4 schools, 576 students	5	Suburban New York School District	Matched on pretests, demographics.	SAT	+0.21
Schreiber, Lomis, & Mys (1988) (WICAT)	Matched post hoc	3 year	Schools: 1 WICAT, 3 control; 254 students	1-4	Schools in Dearborn, Michigan	Matched on ethnicity and Cognitive Abilities Test.	ITBS	+0.21

(continued)

TABLE 2 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup	Overall effect size
Stone (1996) (Exploring Measurement Time and Money)	Matched post hoc	3 years	2 schools; students: 40 CAI, 74 control	2	Middle-class schools	Matched on SES and cognitive abilities tests.	ITBS	+1.16
Borton (1988)	Matched post hoc	1 year	1 school; students: 36 CAI, 56 control	5	Suburban school near San Diego	Matched on pretests, demographics.	CTBS	+0.68
Computer-managed learning systems Accelerated Math								
Ysseldyke & Bolt (2006)	Randomized quasi-experiment	1 year	5 schools, 823 students	2-5	Schools in Texas, Alabama, South Carolina, and Florida	Random assignment, matched on pretests, demographics.	Terra Nova	+0.03

(continued)

TABLE 2 (continued)

Study	Design	Duration	N	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
S. M. Ross & Nunnery (2005)	Matched post hoc	1 year	2,350 Accelerated Math students, 1,841 control students	3-5	Schools in southern Mississippi	Matched on pretests, demographics.	MCT		+0.04
Ysseldyke, Spicuzza, Kosciolek, Teelucksingh, et al. (2003)	Matched post hoc	1 year	Students: 397 Accelerated Math, 913 control	3-5	Schools in large urban district in the Midwest	Matched on pretests.	NALT		+0.11
							Within-class comparison	+0.08	
							District comparison	+0.14	

(continued)

TABLE 2 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Spicuzza et al. (2001)	Matched post hoc	5 months	Students: 137 Accelerated Math, 358 control	4, 5	Large urban district in the Midwest	Matched on pretests, demographics.	NALT		+0.17
							Within-class comparison	+0.19	
							District comparison	+0.14	
Johnson-Scott (2006)	Matched post hoc	1 year	3 schools, 7 classes, 82 students	5	Schools in rural Mississippi	Matched on pretests, demographics.	MCT		+0.23

Note. See Appendix B for the table of abbreviations.

measures of math computations. Across all studies from which effect sizes could be computed, these effects are meaningful in light of the fact that CAI is a supplemental approach, rarely occupying more than three 30-minute sessions weekly (and often alternating with CAI reading instruction). The median effect size was +0.19. This is larger than the median found for the curriculum studies (+0.10), and it is based on many more studies (38 vs. 13) and on many more randomized and randomized quasi-experimental studies (15 vs. 2). However, it is important to note that most of these studies are quite old, and they usually evaluated programs that are no longer commercially available.

In the Slavin et al. (2007) review of middle and high school math programs, the median effect size across 36 qualifying studies of CAI was +0.17, nearly identical to the estimate for elementary CAI studies.

Although outcomes of studies of CAI are highly variable, most studies do find positive effects, and none significantly favored a control group. Although the largest number of studies has involved Jostens, there is not enough high-quality evidence on particular CAI approaches to recommend any one over another, at least based on student outcomes on standardized tests.

In studies that break down their results by subscales, outcomes are usually stronger for computation than for concepts or problem solving. This is not surprising, as CAI is primarily used as a supplement to help children with computation skills. Because of the hierarchical nature of math computation, CAI has a special advantage in this area because of its ability to assess students and provide them with individualized practice on skills that they have the prerequisites to learn but have not yet learned.

Instructional Process Strategies

Many researchers and reformers have sought to improve children's mathematics achievement by giving teachers extensive professional development on the use of instructional process strategies, such as cooperative learning, classroom management, and motivation strategies (see Hill, 2004). Curriculum reforms and CAI also typically include professional development, of course, but the strategies reviewed in this section are primarily characterized by a focus on changing what teachers do with the curriculum they have, not changing the curriculum.

The programs in this section are highly diverse, so they are further divided into seven categories:

1. cooperative learning,
2. cooperative/individualized programs,
3. direct instruction,
4. mastery learning,
5. professional development focused on math content,
6. professional development focused on classroom management and motivation, and
7. supplemental programs.

A total of 36 studies evaluated instructional process programs. Table 3 summarizes characteristics and outcomes of these studies. For discussions of each study, see http://www.bestevidence.org/_images/word_docs/Eff%20progs%20ES%20math%20Version%201.2%20for%20BEE%2002%2009%2007.doc.

(text continues on p. 475)

TABLE 3
Instructional process strategies: Descriptive information and effect sizes for qualifying studies

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
				Cooperative learning					
Classwide Peer Tutoring									
Greenwood, Delquadri, & Hall (1989) and Greenwood, Terry, Utley, & Montagna (1993)	Randomized quasi-experiment	4 years	123 students	1-4	High-poverty school district in metropolitan area of Kansas City	IQ scores and SES were not significantly different between the groups. IQ and pretest achievement served as covariates.	MAT		+0.33
							2-year follow-up	+0.23	
PALS									
Fuchs, Fuchs, Yazdian, & Powell (2002)	Randomized quasi-experiment	16 weeks	20 classes, 323 students	1	Schools in southeastern city	Matched on pretests.	SAT		+0.10
							High achieving	+0.09	
							Average achieving	+0.10	
							Low achieving	+0.14	

(continued)

TABLE 3 (continued)

Study	Design	Duration	N	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Fuchs, Fuchs, & Karns (2001)	Randomized quasi-experiment	15 weeks	20 teachers, 228 students	K	Schools in Southeastern city	Matched on pretests.	SESAT		+0.24
							High achieving	–0.41	
							Average achieving	+0.52	
							Low achieving	+0.51	
							Disability	+0.65	
							SAT		
							High achieving	+0.85	
							Average achieving	–0.20	
							Low achieving	+0.47	
							Disability	+0.20	

(continued)

TABLE 3 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
STAD									
Mevarech (1985)	Random assignment	15 weeks	67 students	5	Israeli school, middle-class students	Random assignment, matched on pretests, demographics.	Objective-based test		+0.19
							Computation	+0.36	
							Comprehension	+0.01	
Glassman (1989)	Randomized quasi-experiment	6 months	2 schools, 24 classes, 441 students	3-5	Schools in diverse, suburban district in Long Island, New York	Random assignment, matched on pretests.	ITBS		+0.01
Mevarech (1991)	Randomized quasi-experiment	12 weeks	54 students	3	Low SES school in Israel	Matched on pretests.	Teacher-designed test		+0.60
							Low achievers	+0.86	
							High achievers	+0.69	

(continued)

TABLE 3 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Suyanto (1998)	Matched	4 months	10 schools, 30 classes, 664 students	3-5	Schools across rural Indonesia	Matched on pretests, demographics.	Indonesian Elementary School Test of Learning		+0.40
Student Team Mastery Learning									
Mevarech (1985)	Random assignment	15 weeks	67 students	5	Israeli school, middle-class students	Random assignment, matched on pretests.	Objective-based test		+0.29
							Computation	+0.36	
							Comprehension	+0.21	
Mevarech (1991)	Randomized quasi-experiment	12 weeks	54 students	3	Low SES school in Israel	Random assignment, matched on pretests.	Teacher-designed test		+0.55
							Low achievers	+0.80	
							High achievers	+0.80	
Cooperative/Individualized Programs									
TAI									
Slavin & Karweit (1985)	Random assignment	16 weeks	17 classes, 382 students	3-5	Hagerstown, Maryland	Matched on pretests.	CTBS		+0.28

(continued)

TABLE 3 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
							TAI vs. MMP		
							Computation	+0.39	
							Concepts and applications	+0.01	
							TAI vs. Control		
							Computation	+0.67	
							Concepts and applications	+0.06	
Slavin & Karweit (1985)	Random assignment	18 weeks	10 classes, 212 students	4-6	Inner-city Wilmington, Delaware		CTBS		+0.38
							Computation	+0.76	
							Concepts and applications	0.00	
Slavin, Madden, & Leavey (1984b)	Matched	24 weeks	59 classes, 1,367 students	3-5	Schools located in middle-class suburb of Baltimore, Maryland	Matched on CAT pretests.	CTBS		+0.14
							Computation	+0.18	
							Concepts and applications	+0.10	
							Students with special needs		
							Computation	+0.19	
							Concepts and applications	+0.23	

(continued)

TABLE 3 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
R. J. Stevens & Slavin (1995)	Matched	2 years	45 classes, 873 students	2-6	Schools located in diverse Baltimore suburb	Matched on pretests, demographics.	CAT		+0.20
							Computation	+0.29	
							Concepts and applications	+0.10	
							Students with special needs		
							Computation	+0.59	
							Concepts and applications	+0.35	
							Gifted students		
							Computation	+0.59	
							Concepts and applications	+0.19	

(continued)

TABLE 3 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Karper & Melnick (1993)	Matched	1 year	165 students	4, 5	School in affluent suburb of Harrisburg, Pennsylvania	Matched on district test.	District standardized test		–.0.11
							Computation	–0.11	
							Concepts	–0.22	
Project CHILD									
Orr (1991)	Matched	3 years	2 schools, 186 students	2-5	Schools in northeast and northwest Florida	Matched on pretests.	Standardized achievement tests		+0.69
Direct instruction									
CMC									
Snider & Crawford (1996)	Random assignment	1 year	1 school, 46 students	4	Rural school in northern Wisconsin	Random assignment, matched on pretests,	NAT		+0.26
							Computation	+0.72	
							Concepts and problem solving	+0.01	

(continued)

TABLE 3 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Tarver & Jung (1995)	Matched	2 years	88 students: 21 CMC, 67 Mathematics Their Way/CGI	1, 2	School in suburban Midwest	Matched on pretests.	CTBS Computation Concepts and Applications	 +2.13 +0.68	+1.33
Crawford & Snider (2000)	Matched post hoc	1 year	1 school, 52 students	4	Rural school in northern Wisconsin	Matched on pretests.	NAT Computation Concepts and problem solving	 +0.33 +0.39	+0.41
User-Friendly Direct Instruction									
Grossen & Ewing (1994)	Randomized	2 years	1 school, 45 students	5, 6	School in Boise, Idaho	Matched on pretests.	Woodcock-Johnson Applications ITBS Concepts Prob. Solving Operations	 +0.50 +0.44 +0.17 +0.97	+0.52

(continued)

TABLE 3 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
				Mastery Learning					
Mastery Learning									
Mevarech (1985)	Random assignment	15 weeks	67 students	5	Israeli school-middle-class students	Random assignment, matched on pretests, demographics.	Objective-Based Test Computation Comprehension	 +0.55 +0.28	+0.42
Mevarech (1991)	Randomized quasi-experiment	12 weeks	85 students	3	Low SES school in Israel	Random assignments, matched on pretests.	Teacher-designed test Low achievers High achievers	 +1.10 +2.20	+1.08
Cox (1986)	Matched	6 months	173 students	5	School in southwest-ern Missouri	Matched on pretests.	ITBS		+0.22
Monger (1989)	Matched post hoc	1 year	140 students	2, 5	Schools located in suburban districts in Oklahoma	Matched on pretests, demographics.	MAT-6 Grade 2 Grade 3	 –0.09 –0.27	–0.18

(continued)

TABLE 3 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Anderson, Scott, & Hutlock (1976)	Matched post hoc	1 year	2 schools	1-6	Schools near Cleveland, Ohio	Matched on ability measures.	CAT		+0.04
							Computation	+0.17	
							Problem solving	+0.07	
							Concepts	–0.12	
Professional development focused on math content									
CGI									
Carpenter, Fennema, Peterson, Chiang, & Loef (1989)	Randomized quasi-experiment	1 year	40 teachers	1	Schools in and around Madison, Wisconsin	Random assignment, matched on pretests.	ITBS		+0.24
							Computation	+0.22	
							Problem solving	+0.25	
Dynamic pedagogy									
Armour-Thomas, Walker, Dixon-Roman, Mejia, & Gordon (2006)	Matched	1 year	2 schools, 120 students	3	Schools in New York suburb	Matched on pretests, demographics.	Terra Nova		+0.32

(continued)

TABLE 3 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Professional development focused on classroom management and motivation									
MMP									
Slavin & Karweit (1985)	Random assignment	16 weeks	22 classes, 366 students	3-5	Schools in and around Hagerstown, Maryland	Random assignment, matched on pretests.	CTBS		+0.57
							MMP two-group	+0.84	
							MMP whole-class	+0.29	
T. L. Good & Grouws (1979)	Randomized quasi-experiment	3 months	27 schools, 40 teachers	4	Tulsa, Oklahoma, school district	Random assignment, matched on pretests, demographics.	SRA		+0.33
Ebmeier & Good (1979)	Randomized quasi-experiment	15 weeks	28 schools, 39 teachers	4	Schools in large South-western school district	Random assignment, matched on pretests.	SRA		+0.42

(continued)

TABLE 3 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup	Overall effect size
CMCD								
Freiberg, Prokosch, Treiser, & Stein (1990)	Matched post hoc	1 to 2 years 1 year 1 year	10 schools, 699 students	2-5	Low-performing, high-minority schools in Houston, Texas	Matched on pretests, demographics.	MAT-6, Texas Educational Assessment of Minimal Skills	+0.40
Freiberg, Connell, & Lorentz (2001)	Matched post hoc		7 schools, 543 students	4-6	Chapter 1 schools in large urban city in the Southwest United States. Students mainly Latino.	Matched on pretests, demographics.	TAAS	+0.33
Opuni (2006)	Matched post hoc		456 students	3	Newark Public Schools	Matched on pretests, demo-graphics.	Stanford-9	+0.53

(continued)

TABLE 3 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
				Supplemental programs					
Small-group tutoring									
Fuchs et al. (2005)	Random assignment	16 weeks	127 students	1	Schools in southeastern metropolitan district. Students at risk.	Matched on pretests, demographics.	Woodcock-Johnson		+0.37
							Calculations	+0.56	
							Applied problems	+0.09	
							Experimenter-made measure		
							Addition	+0.36	
							Subtraction	+0.03	
							Concepts and applications	+0.62	
							Story problems	+0.59	

(continued)

TABLE 3 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristics	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
Every Day Counts									
Great Source (2006)	Matched	6 months	13 schools, 587 students	5	Schools in New Haven, Connecticut	Matched on pretests, demographics.	Assessment modeled after CMT		+0.15
Project SEED									
W. Webster (1995)	Matched post hoc	1-3 semesters	Students: Cycle 1: 732; Cycle 2: 558; Cycle 3: 249	4-6	Detroit Public Schools	Matched on pretests, demographics.	CAT		+0.73
							1 semester SEED exposure		
							Math Total	+0.29	
							Computation	+0.35	
							Concepts	+0.32	
							2 semesters SEED exposure		
							Math Total	+0.60	
							Computation	+0.58	
							Concepts	+0.68	
							3 semesters SEED exposure		
							Math Total	+0.73	
							Computation	+0.74	
							Concepts	+0.62	

(continued)

TABLE 3 (continued)

Study	Design	Duration	*N*	Grade	Sample characteristi	Evidence of initial equality	Effect sizes by posttest and subgroup		Overall effect size
W. J. Webster & Dryden (1998)	Matched post hoc	14+ weeks	604 students	3	Detroit Public Schools	Matched on pretests, demographics.	MAT-6		+0.25
							Math total	+0.25	
							Concepts and problem solving	+0.28	
							Procedures	+0.17	

Note. See Appendix B for the table of abbreviations.

Research on instructional process strategies tends to be of much higher quality than research on mathematics curricula or CAI. Out of 36 studies, 19 used randomized or randomized quasi-experimental designs. Many had small samples and/or short durations, and in some cases there were confounds between treatments and teachers, but even so, these are relatively high-quality studies, most of which were published in peer-reviewed journals. The median effect size for randomized studies was +0.33, and the median was also +0.33 for all studies taken together.

The research on instructional process strategies identified several methods with strong positive outcomes on student achievement. In particular, the evidence supports various forms of cooperative learning. The median effect size across 9 studies of cooperative learning was +0.29 for elementary programs (and +0.38 for middle and high school programs; see Slavin et al., 2007). Particularly positive effects were found for Classwide Peer Tutoring (CWPT) and Peer-Assisted Learning Strategies (PALS), which are pair learning methods, and Student Teams-Achievement Division (STAD) and TAI Math, which use groups of four. Project CHILD, which also uses cooperative learning, was successfully evaluated in 1 study.

Two programs that focus on classroom management and motivation also had strong evidence of effectiveness in large studies. These are the Missouri Mathematics Project, with three large randomized and randomized quasi-experiments with positive outcomes, and Consistency Management & Cooperative Discipline® (CMCD). Positive effects were also seen for Dynamic Pedagogy and Cognitively Guided Instruction (CGI), which focus on helping teachers understand math content and pedagogy. Four small studies supported direct instruction models, Connecting Math Concepts, and User-Friendly Direct Instruction.

Programs that supplemented traditional classroom instruction also had strong positive effects. These include small-group tutoring for struggling first graders and Project SEED, which provides a second math period focused on high-level math concepts.

The research on these instructional process strategies suggests that the key to improving math achievement outcomes is changing the way teachers and students interact in the classroom. It is important to be clear that the well-supported programs are not ones that just provide generic professional development or professional development focusing on mathematics content knowledge. What characterizes the successfully evaluated programs in this section is a focus on how teachers use instructional process strategies, such as using time effectively, keeping children productively engaged, giving children opportunities and incentives to help each other learn, and motivating students to be interested in learning mathematics.

Overall Patterns of Outcomes

Across all categories, 87 studies met the inclusion criteria for this review, of which 36 used randomized or randomized quasi-experimental designs: 13 studies (2 randomized) of mathematics curricula, 38 (11 randomized, 4 randomized quasi-experimental) of CAI, and 36 (9 randomized, 10 randomized quasi-experimental) of instructional process programs. The median effect size was +0.22. The effect size for randomized and randomized quasi-experimental studies (n = 36) was +0.29, and for fully randomized studies (n = 22) it was +0.28, indicating that randomized studies generally produced effects similar to those of matched quasi-experimental studies. Recall that the matched studies had to meet stringent

methodological standards, so the similarity between randomized and matched outcomes reinforces the observation made by Glazerman, Levy, and Myers (2002) and Torgerson (2006) that high-quality studies with well-matched control groups produce outcomes similar to those of randomized experiments.

Overall effect sizes differed, however, by type of program. Median effect sizes for all qualifying studies were +0.10 for mathematics curricula, +0.19 for CAI programs, and +0.33 for instructional process programs. Effect sizes were above the overall median (+0.22) in 15% of studies of mathematics curricula, 37% of CAI studies, and 72% of instructional process programs. The difference in effect sizes between the instructional process and other programs is statistically significant ($\chi^2 = 15.71$, $p < .001$).

With only a few exceptions, effects were similar for disadvantaged and middle-class students and for students of different ethnic backgrounds. Effects were also generally similar on all subscales of math tests, except that CAI and TAI Math generally had stronger effects on measures of computations than on measures of concepts and applications.

Summarizing Evidence of Effectiveness for Current Programs

In several recent reviews of research on outcomes of various educational programs, reviewers have summarized program outcomes using a variety of standards. This is not as straightforward a procedure as might be imagined, as several factors must be balanced (Slavin, 2008). These include the number of studies, the average effect sizes, and the methodological quality of studies.

The problem is that the number of studies of any given program is likely to be small, so simply averaging effect sizes (as in meta-analyses) is likely to overemphasize small, biased, or otherwise flawed studies with large effect sizes. For example, in the present review, there are several very small matched studies with effect sizes in excess of +1.00, and these outcomes cannot be allowed to overbalance large and randomized studies with more modest effects. Emphasizing numbers of studies can similarly favor programs with many small, matched studies, which may collectively be biased toward positive findings by file drawer effects. The difference in findings for CAI programs between small numbers of randomized experiments and large numbers of matched experiments shows the danger of emphasizing numbers of studies without considering quality. Finally, emphasizing methodological factors alone risks eliminating most studies or emphasizing studies that may be randomized but are very small, confounding teacher and treatment effects, or may be brief, artificial, or otherwise not useful for judging the likely practical impact of a treatment.

In this review, we applied a procedure for characterizing the strength of the evidence favoring each program that attempts to balance methodological, replication, and effect size factors. Following the What Works Clearinghouse (2006), the Comprehensive School Reform Quality Center (2005), and Borman, Hewes, Overman, and Brown (2003), but placing a greater emphasis on methodological quality, we categorized programs as follows (see Slavin, 2007):

- *Strong evidence of effectiveness*—at least one large randomized or randomized quasi-experimental study, or two smaller studies, with a median effect size of at least +0.20. A large study is defined as one in which at least 10 classes or schools, or 250 students, were assigned to treatments.

◒ *Moderate evidence of effectiveness*—at least one randomized or randomized quasi-experimental study, or a total of two large or four small qualifying matched studies, with a median effect size of at least +0.20.

◔ *Limited evidence of effectiveness*—at least one qualifying study of any design with a statistically significant effect size of at least +0.10.

○ *Insufficient evidence of effectiveness*—one or more qualifying study of any design with nonsignificant outcomes and a median effect size less than +0.10.

N *No qualifying studies*.

Table 4 summarizes currently available programs falling into each of these categories (programs are listed in alphabetical order within each category). Note that programs that are not currently available, primarily the older CAI programs, do not appear in the table, as it is intended to represent the range of options from which today's educators might choose.

In line with the previous discussions, the programs represented in each category are strikingly different. In the Strong Evidence category appear five instructional process programs, four of which are cooperative learning programs: Classwide Peer Tutoring, PALS, STAD, and TAI Math. The fifth program is a classroom management and motivation model, the Missouri Mathematics Project.

The Moderate Evidence category is also dominated by instructional process programs, including two supplemental designs, small-group tutoring, and Project SEED, as well as CGI, which focuses on training teachers in mathematical concepts, and CMCD, which focuses on school and classroom management and motivation. Connecting Math Concepts, an instructional process program tied to a specific curriculum, also appears in this category. The only current CAI program with this level of evidence is Classworks.

The Limited Evidence category includes five math curricula: Everyday Mathematics, Excel Math, Growing With Mathematics, Houghton Mifflin Mathematics, and Knowing Mathematics. Dynamic Pedagogy, Project CHILD, and Mastery Learning—instructional process programs—are also in this category, along with Lightspan and Accelerated Math. Four programs, listed under the Insufficient Evidence category, had only one or two studies, which failed to find significant differences. The final category, No Qualifying Studies, lists 48 programs.

Discussion

The research reviewed in this article evaluates a broad range of strategies for improving mathematics achievement. Across all topics, the most important conclusion is that there are fewer high-quality studies than one would wish for. Although a total of 87 studies across all programs qualified for inclusion, there were small numbers of studies on each particular program. There were 36 studies, 19 of which involved instructional process strategies, that randomly assigned schools, teachers, or students to treatments, but many of these tended to be small and therefore to confound treatment effects with school and teacher effects. There were several large-scale, multiyear studies, especially of mathematics curricula, but these tended to be post hoc matched quasi-experiments, which can introduce serious selection bias. Clearly, more randomized evaluations of programs used on a significant scale over a year or more are needed.

TABLE 4
Summary of evidence supporting currently available elementary mathematics programs

◐ *Strong evidence of effectiveness*
Classwide Peer Tutoring (IP)
Missouri Mathematics Project (IP)
Peer-Assisted Learning Strategies (IP)
Student Teams–Achievement Divisions (IP)
TAI Math (IP/MC)
◒ *Moderate evidence of effectiveness*
Classworks (CAI)
Cognitively Guided Instruction (IP)
Connecting Math Concepts (IP/MC)
Consistency Management & Cooperative Discipline® (IP)
Project SEED (IP)
Small-group tutoring (IP)
◔ *Limited evidence of effectiveness*
Accelerated Math (CAI)
Dynamic Pedagogy (IP)
Every Day Counts (IP)
Everyday Mathematics (MC)
Excel Math (MC)
Growing With Mathematics (MC)
Houghton Mifflin Mathematics (MC)
Knowing Mathematics (MC)
Lightspan (CAI)
Mastery Learning (IP)
Project CHILD (IP/CAI)
○ *Insufficient evidence*
Math Steps (MC)
Math Trailblazers (MC)
Saxon Math (MC)
Scott Foresman–Addison Wesley (MC)
N *No qualifying studies*
Academy of Math® (CAI)
Adventures of Jasper Woodbury (IP/CAI)
Advanced Learning System (CAI)
AIMSweb® Pro Math (CAI)
Approach & Connect (IP)
Barrett Math Program (IP)
Blast Off Math (IP)
Box It or Bag It (IP)
BoxerMath (CAI)
Breakaway Math (IP)

(continued)

TABLE 4 (continued)

Breakaway Math (IP/MC)
Bridges in Mathematics (MC)
Buckle Down (IP)
Building Math Ideas (IP)
Compass Learning (current version) (CAI)
Connected Tech (CAI)
Corrective Math (MC)
Count, Notice, & Remember (IP)
Destination Math Series (CAI)
Elementary Math With Pizzazz! (IP)
Facts for Life (IP)
Facts that Last (IP)
First in Math® (CAI)
Foundations in Math (IP/MC)
Great Explorations in Math and Science (IP/MC)
Groundworks (IP)
Harcourt Math (MC)
HeartBeeps® (CAI)
Investigations in Number, Data, and Space (MC)
Journey Math (CAI)
JumpStart Math (CAI)
Knowledge Box® (CAI)
Larson's Elementary Math
LearnStar (CAI)
Macmillan McGraw-Hill Math (MC)
Mastery Math (MC/IP)
Math Achievement Predictors (IP)
Math Advantage (MC)
Math Blasters (CAI)
Math Central (MC)
Math Coach (MC/IP)
Math Explorations and Applications (MC)
Math Expressions (MC)
Math to Know (IP)
Math Made Easy (CAI)
Math Matters (IP)
Math & Me Series (MC)
Math & Music (CAI)
Math in My World (MC)
Math for the Real World™ (CAI)
Math Their Way (MC)
MathAmigo (CAI)
Mathematics Plus (MC)

(continued)

TABLE 4 (continued)

Mathematics Their Way (MC)
MathFact (IP)
Mathletics (MC)
MathRealm (CAI)
MathStart® (IP)
MathWings (IP/MC)
McGraw-Hill Mathematics (MC)
Moving with Math® (MC/IP)
New Century Integrated Instructional System (IP/CAI)
New Century Mathematics (MC)
Number Power (MC)
Number Worlds (MC/IP)
Numeracy Recovery
Opening Eyes to Mathematics
Orchard Mathematics (CAI)
PLATO® (CAI)
Problem Solving Step by Step (IP/MC)
Progress in Mathematics (MC)
Project IMPACT (IP)
Project M3: Mentoring Mathematical Minds (MC)
Rational Number Project (MC)
Real Math (MC)
Reciprocal Peer Tutoring (IP)
Scott Foresman Math Around the Clock (IP/MC)
Singapore Math (MC)
SkillsTutor/CornerStone2 (CAI)
Strategic Math Series (IP)
Strength in Numbers (IP)
SuccessMaker (CAI) (Current version)
Thinking Mathemementally (IP)
Time4Learning (MC)
TIPS Math (IP)
TouchMath (IP)
Visual Mathematics
Voyages (IP/MC)
Waterford Early Math (CAI)
Yearly Progress Pro (CAI)

Note. IP = instructional process strategies; MC = mathematics curricula; CAI = computer-assisted instruction.

This being said, there were several interesting patterns in the research on elementary mathematics programs. One surprising observation is the lack of evidence that it matters very much which textbook schools choose (median ES = +0.10 across 13 studies). Quality research is particularly lacking in this area, but the mostly matched

post hoc studies that do exist find modest differences between programs. NSF-funded curricula such as Everyday Mathematics, Investigations, and Math Trailblazers might have been expected to at least show significant evidence of effectiveness for outcomes such as problem solving or concepts and applications, but the quasi-experimental studies that qualified for this review find little evidence of strong effects even in these areas. The large national study of these programs by Sconiers et al. (2003) found effect sizes of only +0.10 for all outcomes, and the median effect size for 5 studies of NSF-funded programs was +0.12.

It is possible that the state assessments used in the Sconiers et al. (2003) study and other studies may have failed to detect some of the more sophisticated skills taught in NSF-funded programs but not other programs, a concern expressed by Schoenfeld (2006) in his criticism of the What Works Clearinghouse. However, in light of the small effects seen on outcomes such as problem solving, probability and statistics, geometry, and algebra, it seems unlikely that misalignment between the NSF-sponsored curricula and the state tests accounts for the modest outcomes.

Studies of CAI found a median effect size (ES = +0.19) higher than that found for mathematics curricula, and there were many more high-quality studies of CAI. A number of studies showed substantial positive effects of using CAI strategies, especially for computation, across many types of programs. However, the highest quality studies, including the few randomized experiments, mostly found no significant differences.

CAI effects in math, although modest in median effect size, are important in light of the fact that in most studies CAI was used for only about 30 minutes three times a week or less. The conclusion that CAI is effective in math is in accord with the findings of a recent review of research on technology applications by Kulik (2003), who found positive effects of CAI in math but not in reading.

The most striking conclusion from the review, however, is the evidence supporting various instructional process strategies. Twenty randomized experiments and randomized quasi-experiments found impressive affects (median ES = +0.33) for programs that target teachers' instructional behaviors rather than math content alone. Several categories of programs were particularly supported by high-quality research. Cooperative learning methods, in which students work in pairs or small teams and are rewarded based on the learning of all team members, were found to be effective in 9 well-designed studies, 8 of which used random assignment, with a median effect size of +0.29. These included studies of Classwide Peer Tutoring, PALS, and STAD. Team Accelerated Instruction, which combines cooperative learning and individualization, also had strong evidence of effectiveness. Another well-supported approach included programs that focus on improving teachers' skills in classroom management, motivation, and effective use of time, in particular the Missouri Mathematics Project and CMCD. Studies supported programs focusing on helping teachers introduce mathematics concepts effectively, such as CGI, Dynamic Pedagogy, and Connecting Math Concepts.

Supplementing classroom instruction with well-targeted supplementary instruction is another strategy with strong evidence of effectiveness. In particular, small-group tutoring for first graders struggling in math and Project SEED, which provides an additional period of instruction from professional mathematicians, have strong evidence.

The debate about mathematics reform has focused primarily on curriculum, not on professional development or instruction (see, e.g., American Association for the Advancement of Science, 2000; NRC, 2004). Yet this review suggests that in terms of outcomes on traditional measures, such as standardized tests and state accountability assessments, curriculum differences appear to be less consequential than instructional differences are. This is not to say that curriculum is unimportant. There is no point in teaching the wrong mathematics. The research on the NSF-supported curricula is at least comforting in showing that reform-oriented curricula are no less effective than traditional curricula on traditional measures, and they may be somewhat more effective, so their contribution to nontraditional outcomes does not detract from traditional ones. The movement led by the National Council of Teachers of Mathematics to focus math instruction more on problem solving and concepts may account for the gains over time on NAEP, which itself focuses substantially on these domains.

Also, it is important to note that the three types of approaches to mathematics instruction reviewed here do not conflict with each other and may have additive effects if used together. For example, schools might use an NSF-supported curriculum such as Everyday Mathematics with well-structured cooperative learning and supplemental CAI, and the effects may be greater than those of any of these programs by themselves. However, the findings of this review suggest that educators and researchers might do well to focus more on how mathematics is taught, rather than expecting that choosing one or another textbook by itself will move their students forward.

As noted earlier, the most important problem in mathematics education is the gap in performance between middle- and lower-class students and between White and Asian American students and African American, Hispanic, and Native American students. The studies summarized in this review took place in widely diverse settings, and several of them reported outcomes separately for various subgroups. Overall, there is no clear pattern of differential effects for students of different social class or ethnic background. Programs found to be effective with any subgroup tend to be effective with all groups. Rather than expecting to find programs with different effects on students in the same schools and classrooms, the information on effective mathematics programs might better be used to address the achievement gap by providing research-proven programs to schools serving many disadvantaged and minority students. Federal Reading First and Comprehensive School Reform programs were intended to provide special funding to help high-poverty, low-achieving schools adopt proven programs. A similar strategy in mathematics could help schools with many students struggling in math to implement innovative programs with strong evidence of effectiveness, as long as the schools agree to participate in the full professional development process used in successful studies and to implement all aspects of the program with quality and integrity.

The mathematics performance of America's students does not justify complacency. In particular, schools serving many students at risk need more effective programs. This article provides a starting place in determining which programs have the strongest evidence bases today. Hopefully, higher quality evaluations of a broader range of programs will appear in the coming years. What is important is that we use what we know now at the same time that we work to improve our knowledge base in the future so that our children receive the most effective mathematics instruction we can give them.

APPENDIX A
Studies not included in the review

Author	Cited by	Reason not included and comments
	Mathematics Curricula	
Bridges in Mathematics		
Math Learning Center (2003)		No adequate control group
Everyday Mathematics		
Briars (2004)		No adequate control group
Briars & Resnick (2000)	NRC	No adequate control group
Carroll (1993)	NRC	No adequate control group
Carroll (1994-1995)	NRC	No adequate control group
Carroll (1995)	NRC	Inadequate outcome measure
Carroll (1996a)	NRC	No adequate control group
Carroll (1996b)	NRC	Inadequate outcome measure
Carroll (1996c)	NRC	No adequate control group
Carroll (1997)		Insufficient match, no pretest
Carroll (1998)	NRC	Inadequate outcome measure
Carroll (2000)	NRC	Baseline equivalence not established
Carroll (2001a)	NRC	No adequate control group
Carroll (2001b)	NRC	Insufficient match, no pretest
Carroll & Fuson (1998)	NRC	No adequate control group
Carroll & Isaacs (2003)	NRC	Baseline equivalence not established
Carroll & Porter (1994)	NRC	Insufficient match
Drueck, Fuson, Carroll, & Bell (1995)	NRC	No adequate control group
Fuson & Carroll (n.d.-a)	NRC	No adequate control group
Fuson & Carroll (n.d.-b)	NRC	Insufficient match, no pretest
Fuson, Carroll, & Drueck (2000)	NRC	No adequate control group
Mathematics Evaluation Committee (1997)	NRC	Insufficient match
McCabe (2001)		Insufficient match, no pretest
L. A. Murphy (1998)		No adequate control group
Salvo (2005)		Duration less than 12 weeks
Houghton Mifflin Mathematics		
EDSTAR (2004)		Insufficient data
Mehrens & Phillips (1986)		No adequate control group
Sheffield (2004)		No adequate control group
Sheffield (2005)		No adequate control group
Investigations in Number, Data, and Space		
Austin Independent School District (2001)	NRC	Insufficient match, pretest differences not accounted for
Flowers (1998)		Insufficient match and outcome measure

(continued)

APPENDIX A (continued)

Author	Cited by	Reason not included and comments
Gatti (2004a)		Pretest equivalence not documented
Goodrow (1998)	NRC	Insufficient match and outcome measure
McCormick (2005)		Measure inherent to treatment
Mokros, Berle-Carmen, Rubin, & O'Neil (1996)	NRC	Insufficient match and outcome measure
Mokros, Berle-Carmen, Rubin, & Wright (1994)	NRC	Inadequate outcome measure, pretest differences not accounted for
L. G. Ross (2003)		No adequate control group
Math Their Way		
Mayo (1995)		Pretest equivalence not established
McKernan (1992)		Insufficient data
Shawkey (1989)		Insufficient match
Math Trailblazers		
Carter et al. (2003)		No adequate comparison groups, pretest differences not accounted for
Lykens (2003)		No adequate control group
Mathematics Plus		
Rust (1999)		Pretest equivalence not established
MathWings		
Madden, Slavin, & Simons (1997)		No adequate control group
Madden, Slavin, & Simons (1999)		No adequate control group
Number Power		
Cooperative Mathematics Project (1996)		Inadequate outcome measure
Progress in Mathematics		
Beck Evaluation & Testing Associates (2006)		Inadequate control group
Rational Number Project		
Cramer, Post, & delMas (2002)		Duration less than 12 weeks, inadequate outcome measure
Moss & Case (1999)		Test inherent to measure
Real Math (Explorations and Applications)		
Dilworth & Warren (1980)		Inadequate outcome measure, no adequate control group
Rightstart/Number Worlds		
Griffin, Case, & Siegler (1994)		No adequate outcome measure
Saxon Math		
Atkeison-Cherry (2004)		Duration less than 12 weeks
Bolser & Gilman (2003)		No adequate control group
Calvery, Bell, & Wheeler (1993)	NRC	Insufficient match
Fahsl (2001)		

(continued)

APPENDIX A (continued)

Author	Cited by	Reason not included and comments
K. Good, Bickel, & Howley (2006)		Insufficient match, posttest only Pretest equivalence not established
Hansen & Greene (n.d.)	NRC	Insufficient information, no adjusting at posttest
Nguyen (1994)		Insufficient information, no adjusting at posttest
Nguyen & Elam (1993)		Insufficient information
Resendez, Sridharan, & Azin (2006)		Pretest equivalence was not established
Scott Foresman-Addison Wesley		
Gatti (2004b)		Pretest equivalence not established
Simpson (2001)		No adequate control group
Singapore Math		
Ginsburg, Leinwand, Anstrom, & Pollok, E. (2005).		No adequate control group, initial equivalence not established
	Computer-Assisted Instruction	
Accelerated Math		
Atkins (2005)		Pretest equivalence not established
Boys (2003)		Pretest equivalence not established
Brem (2003)		Inadequate outcome measure, no adequate control group
Holmes & Brown (2003)		No adequate control group
Kosciolek (2003)		No adequate control group
Leffler (2001)		No adequate control group
Teelucksingh, Ysseldyke, Spicuzza, & Ginsburg-Block (2001)		Pretest equivalence not established
Ysseldyke, Betts, Thill, & Hannigan (2004)		Large pretest differences
Ysseldyke, Spicuzza, Kosciolek, & Boys (2003)		Insufficient match, pretest differences too large
Ysseldyke, Spicuzza, Kosciolek, Teelucksingh, et al. (2003)		Inadequate outcome measure
Ysseldyke & Tardrew (2002)		Inadequate outcome measure
Ysseldyke & Tardrew (2005)		Inadequate outcome measure
Ysseldyke, Tardrew, Betts, Thill, & Hannigan (2004)		Inadequate outcome measure

(continued)

APPENDIX A (continued)

Author	Cited by	Reason not included and comments
CCC/SuccessMaker		
Crenshaw (1982)		No adequate control group
Donnelly (2004)		Insufficient match, no adjusting at posttest
Kirk (2003)		No adequate control group
Laub & Wildasin (1998)		No adequate control group
McWhirt, Mentavlos, Rose-Baele, & Donnelly (2003)		No adequate control group
Phillips (2001)		Inadequate outcome measure
Tingey & Simon (2001)		No adequate control group
Tingey & Thrall (2000)		No adequate control group
Tuscher (1998)		No adequate control group
Underwood, Cavendish, Dowling, Fogelman, & Lawson (1996)	Kulik (2003) (SRI)	No evidence of pretest equivalence
Wildasin (1994)		No adequate control group
Jostens Learning/Compass Learning		
Brandt & Hutchinson (2005)		No adequate control group
Clariana (1996)	Kulik (2003) (SRI)	Insufficient information provided
Interactive (2003)		No adequate control group
Jamison (2000)		Duration less than 12 weeks
Leiker (1993)	Kulik (2003) (SRI)	Treatment and control used different pretests
Mann, Shakeshaft, Becker, & Kottkamp (1999)		No adequate control group
Moody (1994)		No adequate control group
Rader (1996)		Duration less than 12 weeks
Roy (1993)	Kulik (2003) (SRI)	Insufficient information provided
Sinkis (1993)	Kulik (2003) (SRI)	Insufficient match
J. W. Stevens (1991)	Kulik (2003) (SRI)	Pretest differences too large
Taylor (1999)		No adequate control group
Lightspan/Plato		
Giancola (2000)		No adequate control group
Gwaltney (2000)		Treatment confounded with other programs

(continued)

APPENDIX A (continued)

Author	Cited by	Reason not included and comments
Interactive (2000)		No adequate control group
Interactive (2001)		Program began before pretest
Interactive & Metis Associates (2002)		Program began before pretest
Quinn & Quinn (2001a)		No adequate control group
Quinn & Quinn (2001b)		No adequate control group
Other CAI		
Anelli (1977)		No untreated control group
Axelrod, McGregor, Sherman, & Hamlet (1987)		No adequate control group, duration less than 12 weeks
Bedell (1998)		No adequate control group
Brown & Boshamer (2000) (Fundamentally Math)		Pretest equivalence not demonstrated
Carrier, Post, & Heck (1985)		Inadequate outcome measure
Chang, Sung, & Lin (2006)		Duration less than 12 weeks
Chiang (1978)		Insufficient match
Cognition and Technology Group at Vanderbilt (1992)		Inadequate outcome measure
Dahn (1992) (Wasach)		No evidence of initial equivalence
Dobbins (1993) (Math Concepts and Skills)		No adequate control group
Emihovich & Miller (1988)		Duration less than 12 weeks
Faykus (1993) (WICAT)		Duration less than 12 weeks
Foley (1994)		Insufficient sample
Hativa (1998)		Insufficient sample
Haynie (1989)		No adequate control group
Isbell (1993)		No adequate control group
Kastre (1995)		Duration less than 12 weeks
Lin, Podell, & Tournaki-Rein (1994)		Duration less than 12 weeks
McDermott & Watkins (1983)		Insufficient data
Mevarech & Rich (1985)		No accounting for pretest differences
Mills (1997)		No adequate control group
Orabuchi (1992)		No accounting for pretest differences
Perkins (1987)		Duration less than 12 weeks
Podell, Tournaki-Rein, & Lin (1992)		Duration less than 12 weeks
Shiah, Mastropieri, Scruggs, & Fulk (1994-1995)		Inadequate outcome measure
Snow (1993)		No adequate control group
Sullivan (1989)		No adequate control group
Suppes, Fletcher, Zanotti, Lorton, & Searle (1973)		No adequate control group

(continued)

APPENDIX A (continued)

Author	Cited by	Reason not included and comments
Trautman & Howe (2004)		No adequate control group
Trautman & Klemp (2004)		No adequate control group
Vogel, Greenwood-Ericksen, Cannon-Bowers, & Bowers (2006)		Duration less than 12 weeks
Wenglinsky (1998)		No adequate control group
Wodarz (1994)		Pretest equivalence not demonstrated
Instructional Process Strategies		
CGI		
Fennema et al. (1996)		No adequate control group
Villasenor & Kepner (1993)		Insufficient match
Classwide peer tutoring		
Greenwood et al. (1984)		No adequate control group, inadequate outcome measure
DuPaul, Ervin, Hook, & McGoey (1998)		No adequate control group, inadequate outcome measure
CMCD		
Freiberg, Stein, & Huang (1995)		Subset of another study
Freiberg, Huzinec, & Borders (2006)		No adequate control group (artificial control group)
Cooperative learning		
Al-Halal (2001)		Duration less than 12 weeks
Bosfield (2004)		Insufficient data
Brush (1997)		Duration less than 12 weeks
De Russe (1999)		No adequate control group
Gabbert, Johnson, & Johnson (1986)		Duration less than 12 weeks
Gilbert-Macmillan (1983)		Duration less than 12 weeks
Goldberg (1989) (TGT)		Inadequate outcome measure
Hallmark (1994)		Duration less than 12 weeks
D. W. Johnson, Johnson, & Scott (1978)		Duration less than 12 weeks
L. C. Johnson (1985) (Groups of Four)		Greater than 1/2 SD apart at pretest
Lucker, Rosenfield, Sikes, & Aronson (1976) (Jigsaw)		Duration less than 12 weeks
Madden & Slavin (1983)		Duration less than 12 weeks
Martin (1986) (TGT)		Duration less than 12 weeks
Morgan (1994)		Duration less than 12 weeks

(continued)

APPENDIX A (continued)

Author	Cited by	Reason not included and comments
Moskowitz, Malvin, Schaeffer, & Schaps (1983) (Jigsaw)		Outcome measure not achievement based
Nattiv (1994)		Duration less than 12 weeks
Peterson, Janicki, & Swing (1981) (small-group instruction)		Duration less than 12 weeks
Swing & Peterson (1982) (small-group instruction)		Duration less than 12 weeks
Tieso (2005)		Duration less than 12 weeks
Williams (2005)		No adequate outcome measure
Xin (1999)		Treatment confounded with other factors
Zuber (1992)		Duration less than 12 weeks
Curriculum-based measurement		
Allinder & Oats (1997)		No adequate control group, inadequate outcome measure
Clarke & Shinn (2004)		No adequate control group
Fuchs, Fuchs, & Hamlett (1989)		Pretest differences too large
Fuchs, Fuchs, Hamlett, Phillips, & Bentz (1994)		Measure inherent to treatment
Fuchs, Fuchs, Hamlett, & Stecker (1991)		Measure inherent to treatment
Stecker & Fuchs (2000)		No adequate control group
Tsuei (2005)		Inadequate comparison group
Direct instruction-CMC, DISTAR Arithmetic I/II, Corrective Mathematics		
W. C. Becker & Gersten (1982)		Insufficient match
Bereiter & Kurland (1981-1982)		Pretest equivalence not established
Brent & DiObilda (1993)		No accounting for pretest scores
Mac Iver, Kemper, & Stringfield (2003)		Insufficient data
Merrell (1996)		No adequate control group
Meyer (1984)		Insufficient data
Vreeland et al. (1994)		Insufficient match
Wellington, J. (1994)		Inadequate outcome measure
Wilson & Sindelar (1991)		Duration less than 12 weeks
Mastery Learning		
Burke (1980)		Duration less than 12 weeks
Cabezon (1984)		No accounting for pretest differences

(continued)

APPENDIX A (continued)

Author	Cited by	Reason not included and comments
Chan, Cole, & Cahill (1988)		Duration less than 12 weeks
Earnheart (1989)		Duration less than 12 weeks
Gallagher (1991)		No adequate control group
Kersh (1970)		No adequate control group, greater than 1/2 *SD* apart at pretest
Long (1991)		Greater than 1/2 *SD* apart at pretest
Peer-assisted learning		
Fuchs, Fuchs, Phillips, Hamlett, & Karns (1995)		Test inherent to treatment
Fuchs, Fuchs, Hamlett, et al. (1997)		Test inherent to treatment
Project CHILD		
Butzin (2001)		Pretest equivalence not demonstrated
Florida TaxWatch (2005)		Pretest equivalence not demonstrated
Gill (1995)		Pretest equivalence not demonstrated
King & Butzin (1992)		Pretest equivalence not demonstrated
Kromhout (1993)		Pretest equivalence not demonstrated
Kromhout & Butzin (1993)		Pretest equivalence not demonstrated
Project SEED		
W. Webster & Chadbourn (1996)		Treatment confounded with other factors
W. J. Webster (1998)		Not enough information, no pretest information
W. J. Webster & Chadbourn (1989)		Treatment confounded with other factors
W. J. Webster & Chadbourn (1992)		Treatment confounded with other factors
Reciprocal peer tutoring		
Fantuzzo, Davis, & Ginsburg (1995)		Duration less than 12 weeks
Fantuzzo, King, & Heller (1992)		Inadequate control group
Fantuzzo, Polite, & Grayson (1990)		Uneven attrition, duration less than 12 weeks
Ginsburg-Block (1998)		Duration less than 12 weeks
Ginsburg-Block & Fantuzzo (1997)		Duration less than 12 weeks
Heller & Fantuzzo (1993)		Test inherent to treatment
Pigott, Fantuzzo, & Clement (1986)		Duration less than 12 weeks
Schema-based instruction		
Fuchs, Fuchs, Finelli, Courey, & Hamlett (2004)		Inadequate outcome measure

(continued)

APPENDIX A (continued)

Author	Cited by	Reason not included and comments
Fuchs, Fuchs, Finelli, et al. (2006)		Inadequate outcome measure
Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, Hosp, et al. (2003)		Inadequate outcome measure
Fuchs, Fuchs, Prentice, et al. (2004)		Inadequate outcome measure
Jitendra et al. (1998)		No adequate control group
Jitendra & Hoff (1996)		Duration less than 12 weeks
STAD		
Vaughan (2002)		No adequate control group
TAI		
Bryant (1981)		Duration less than 12 weeks
Slavin, Leavey, & Madden (1984)		Duration less than 12 weeks
Slavin, Madden, & Leavey (1984a)		Duration less than 12 weeks
Other instructional process strategies		
Ai (2002)		No adequate control group
Beirne-Smith (1991) (peer tutoring)		Duration less than 12 weeks
Burkhouse, Loftus, Sadowski, & Buzad (2003) (Thinking Mathematics professional development)		No adequate control group
Burton (2005)		Pretest equivalence not established
Campbell, Rowan, & Cheng (1995) (Project IMPACT)		Inadequate outcome measure
Cardelle-Elawar (1990) (metacognition)		Duration less than 12 weeks
Cardelle-Elawar (1992, 1995) (metacognition)		Inadequate outcome measure
Cobb et al. (1991) (problem-centered instructional approach)		Pretest equivalence not established
Craig & Cairo (2005) (QUILT)		No adequate control group
Dev, Doyle, & Valenta (2002) (TouchMath)		No comparison group
Fischer (1990) (part-part-whole curriculum)		Duration less than 12 weeks
Follmer (2001)		Duration less than 12 weeks

(continued)

APPENDIX A (continued)

Author	Cited by	Reason not included and comments
Fuchs, Fuchs, Hamlett, & Appleton (2002)		Inadequate outcome measure
Fuchs, Fuchs, Karns, Hamlett, & Katzaroff (1999) (performance- and assessment-driven instruction)		Inadequate outcome measure
Fuchs, Fuchs, Karns, et al. (1997) (task-focused goals treatment)		Inadequate outcome measure
Fuchs, Fuchs, & Prentice (2004) (problem-solving treatment)		Inadequate outcome measure
Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, Hosp, et al. (2003) (explicitly teaching for transfer)		Inadequate outcome measure
Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, & Schroeter (2003) (self-regulated learning strategies)		Inadequate outcome measure
Fueyo & Bushell (1998) (number line procedures and peer tutoring)		Duration less than 12 weeks
Ginsburg-Block & Fantuzzo (1998) (NCTM standards-based intervention)		Duration less than 12 weeks
Hickey, Moore, & Pellegrino (2001) (Adventures of Jasper Woodbury)		Inadequate outcome measures
Hiebert & Wearne (1993)		Inadequate outcome measure
Hohn & Frey (2002) (SOLVED)		Duration less than 12 weeks
Hooper (1992)		Duration less than 12 weeks
Kopecky (2005) (Math Matters)		No adequate control group
Mason & Good (1993) (MMP, two-group and whole-class teaching)		Measure inherent to treatment, no controlling for pretests
Mercer & Miller (1992) (Strategic Math Series)		No comparison group
New Century Education Corporation (2003) (New Century Integrated Instructional System)		Pretest equivalence not established

(continued)

APPENDIX A (continued)

Author	Cited by	Reason not included and comments
Pellegrino, Hickey, Heath, Rewey, & Vye (1992) (Adventures of Jasper Woodbury)		
Pratton & Hales (1986) (active participation)		Duration less than 12 weeks
Ruffin, Taylor, & Butts (1991) (Barrett Math Program)		Pretest equivalence not established
Shaughnessy & Davis (1998) (Opening Eyes to Mathematics by the Math Learning Center)		No adequate control group
Sherwood (1991) (Adventures of Jasper Woodbury)		Pretest equivalence not established
Sloan (1993) (direct instruction)		Pretest equivalence not established
Stallings (1985)		Insufficient information
Stallings & Krasavage (1986) (Madeline Hunter Model)		Pretest equivalence not established
Stallings, Robbins, Presbrey, & Scott (1986) (Madeline Hunter Model)		Insufficient information
White (1996) (TIPS Math)		Duration less than 12 weeks
Yager, Johnson, Johnson, & Snider (1986) (cooperative learning with group processing)		Duration less than 12 weeks

APPENDIX B
Table of abbreviations

ANCOVA—analysis of covariance
CAI—computer-assisted instruction
CAT—California Achievement Test
CCC—Computer Curriculum Corporation
CGI—cognitively guided instruction
CMC—Connecting Math Concepts
CMCD—Consistency Management & Cooperative Discipline®
CMT—Connecticut Mastery Test
CTBS—Comprehensive Test of Basic Skills
ERIC—Education Resources Information Center

(continued)

APPENDIX B (continued)

ISAT—Illinois Standards Achievement Test
ISTEP—Indiana Statewide Testing for Educational Progress
ITBS—Iowa Test of Basic Skills
MAT—Metropolitan Achievement Test
MCAS—Massachusetts Comprehensive Assessment System
MCT—Mississippi Curriculum Test
MMP—Missouri Mathematics Project
NAEP—National Assessment of Educational Progress
NALT—Northwest Achievement Level Test
NAT—National Achievement Test
NCTM—National Council of Teachers of Mathematics
NRC—National Research Council
NSF—National Science Foundation
PALS—peer-assisted learning strategies
SAT—Stanford Achievement Test
SES—socioeconomic status
SESAT—Stanford Early School Achievement Test
SRA—Science Research Associates
SRI—Scholastic Reading Inventory
STAD—student teams–achievement division
TAAS—Texas Assessment of Academic Skills
TAI—team-assisted individualization
TCAP—Tennessee Comprehensive Achievement Test
TGT— Teams-games-tournaments
WASL—Washington Assessment of Student Learning
WICAT—Wireless Internet Center for Advanced Technology

References

Abram, S. L. (1984). *The effect of computer assisted instruction on first grade phonics and mathematics achievement computation*. Unpublished doctoral dissertation, Northern Arizona University.

Ai, X. (2002). *District Mathematics Plan evaluation: 2001–2002 evaluation report.* Los Angeles: Program Evaluation and Research Branch, Los Angeles Unified School District.

Al-Halal, A. (2001). *The effects of individualistic learning and cooperative learning strategies on elementary students' mathematics achievement and use of social skills*. Unpublished doctoral dissertation, Ohio University.

Alifrangis, C. M. (1991). An integrated learning system in an elementary school: Implementation, attitudes, and results. *Journal of Computing in Childhood Education, 2*(3), 51–66.

Allinder, R. M., & Oats, R. G. (1997). Effects of acceptability on teachers' implementation of curriculum-based measurement and student achievement in mathematics computation. *Remedial and Special Education, 18*(2), 113–120.

American Association for the Advancement of Science. (2000). *Middle school mathematics textbooks: A benchmark-based evaluation*. Washington, DC: Author.

Anderson, L., Scott, C., & Hutlock, N. (1976, April). *The effects of a mastery learning program on selected cognitive, affective and ecological variables in Grades 1 through 6*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Anelli, C. (1977). *Computer-assisted instruction and reading achievement of urban third and fourth graders*. Unpublished doctoral dissertation, Rutgers University.

Armour-Thomas, E., Walker, E., Dixon-Roman, E. J., Mejia, B., & Gordon, E. J. (2006, April). *An evaluation of dynamic pedagogy on third-grade math achievement of ethnic minority children*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Atkeison-Cherry, N. K. (2004). *A comparative study of mathematics learning of third-grade children using a scripted curriculum and third-grade children using a non-scripted curriculum*. Unpublished doctoral dissertation, Union University.

Atkins, J. (2005). *The association between the use of Accelerated Math and students' math achievement*. Unpublished doctoral dissertation, East Tennessee State University.

Austin Independent School District. (2001). *Austin collaborative for mathematics education, 1999–2000 evaluation*. Unpublished manuscript.

Axelrod, S., McGregor, G., Sherman, J., & Hamlet, C. (1987). Effects of video games as reinforcers for computerized addition performance. *Journal of Special Education Technology, 9*(1), 1–8.

Baker, S., Gersten, R., & Lee, D. (2002). A synthesis of empirical research teaching mathematics to low-achieving students. *Elementary School Journal, 103*(1), 51–69.

Bass, G., Ries, R., & Sharpe, W. (1986). Teaching basic skills though microcomputer assisted instruction. *Journal of Educational Computing Research, 2*(2), 207–219.

Beck Evaluation & Testing Associates. (2006). *Progress in mathematics 2006: Grade 1 pre–post field test evaluation*. Pleasantville, NY: Author.

Becker, H. J. (1992). Computer-based integrated learning systems in the elementary and middle grades: A critical review and synthesis of evaluation reports. *Journal of Educational Computing Research, 8*(1), 1–41.

Becker, H. J. (1994). Mindless or mindful use of integrating learning systems. *International Journal of Educational Research, 21*(1), 65–79.

Becker, W. C., & Gersten, R. (1982). A follow-up of Follow Through: The later effects of the direct instruction model on children in the fifth and sixth grades. *American Educational Research Journal, 19*(1), 75–92.

Bedell, J. P. (1998). *Effects of reading and mathematics software formats on elementary students' achievement*. Unpublished doctoral dissertation, University of Miami.

Beirne-Smith, M. (1991). Peer tutoring in arithmetic for children with learning disabilities. *Exceptional Children, 57,* 330–337.

Bereiter, C., & Kurland, M. (1981-1982). A constructive look at Follow-Through results. *Interchange, 12,* 1–22.

Birch, J. H. (2002). *The effects of the Delaware Challenge Grant Program on the standardized reading and mathematics test scores of second and third grade students in the Caesar Rodney School District*. Unpublished doctoral dissertation, Wilmington College.

Biscoe, B., & Harris, B. (2004). *Growing With Mathematics: National evaluation study summary report*. DeSoto, TX: Wright Group.

Bolser, S., & Gilman, D. (2003). *Saxon Math, Southeast Fountain Elementary School: Effective or ineffective?* (ERIC Document Reproduction Service No. ED474537).

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*(2), 125–230.

Borton, W. M. (1988). The effects of computer managed mastery learning on mathematics test scores in the elementary school. *Journal of Computer-Based Instruction, 15*(3), 95–98.

Bosfield, G. (2004). *A comparison of traditional mathematical learning and cooperative mathematical learning*. Unpublished master's thesis, California State University, Dominguez Hills.

Boys, C. J. (2003). *Mastery orientation through task-focused goals: Effects on achievement and motivation*. Unpublished master's thesis, University of Minnesota.

Brandt, W. C., & Hutchinson, C. (2005). *Romulus Community Schools comprehensive school reform evaluation.* Naperville, IL: Learning Point Associates.

Brem, S. K. (2003). *AM users outperform controls when exposure and quality of interaction are high: A two-year study of the effects of Accelerated Math and Math Renaissance performance in a Title I elementary school* [tech. rep.]. Tempe: Arizona State University.

Brent, G., & DiObilda, N. (1993). Effects of curriculum alignment versus direct instruction on urban children. *Journal of Educational Research, 86*(6), 333–338.

Briars, D., & Resnick, L. (2000). *Standards, assessments—and what else? The essential elements of standards-based school improvement.* Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, UCLA.

Briars, D. J. (2004, July). *The Pittsburgh story: Successes and challenges in implementing standards-based mathematics programs.* Paper presented at the meeting of the UCSMP Everyday Mathematics Leadership Institute, Lisle, IL.

Brown, F., & Boshamer, C. C. (2000). Using computer assisted instruction to teach mathematics: A study. *Journal of the National Alliance of Black School Educators, 4*(1), 62–71.

Brush, T. (1997). The effects on student achievement and attitudes when using integrated learning systems with cooperative pairs. *Educational Research and Development, 45*(1), 51–64.

Bryant, R. R. (1981). *Effects of team-assisted individualization on the attitudes and achievement of third, fourth, and fifth grade students of mathematics*. Unpublished doctoral dissertation, University of Maryland.

Burke, A. J. (1980). *Students' potential for learning contrasted under tutorial and group approaches to instruction*. Unpublished doctoral dissertation, University of Chicago.

Burkhouse, B., Loftus, M., Sadowski, B., & Buzad, K. (2003). *Thinking Mathematics as professional development: Teacher perceptions and student achievement.* Scranton, Pennsylvania: Marywood University. (ERIC Document Reproduction Service No. ED482911)

Burton, S. K. (2005). *The effects of ability grouping on academic gains of rural elementary school students*. Unpublished doctoral dissertation, Union University.

Butzin, S. M. (2001). Using instructional technology in transformed learning environments: An evaluation of Project CHILD. *Journal of Research on Computing in Education, 33,* 367–373.

Cabezon, E. (1984). *The effects of marked changes in student achievement pattern on the students, their teachers, and their parents: The Chilean case.* Unpublished doctoral dissertation, University of Chicago.

Calvery, R., Bell, D., & Wheeler, G. (1993, November). *A comparison of selected second and third graders' math achievement: Saxon vs. Holt*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, New Orleans, LA.

Campbell, P. F., Rowan, T. E., & Cheng, Y. (1995, April). *Project IMPACT: Mathematics achievement in predominantly minority elementary classrooms attempting reform.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Cardelle-Elawar, M. (1990). Effects of feedback tailored to bilingual students' mathematics needs on verbal problem solving. *Elementary School Journal, 91,* 165–175.

Cardelle-Elawar, M. (1992). Effects of teaching metacognitive skills to students with low mathematics ability. *Teaching and Teacher Education, 8,* 109–121.

Cardelle-Elawar, M. (1995). Effects of metacognitive instruction on low achievers in mathematics problems. *Teaching and Teacher Education, 11,* 81–95.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Education Research Journal, 26,* 499–531.

Carrier, C., Post, T., & Heck, W. (1985). Using microcomputers with fourth-grade students to reinforce arithmetic skills. *Journal for Research in Mathematics Education, 16*(1), 45–51.

Carroll, W. (1995). *Report on the field test of fifth grade Everyday Mathematics*. Chicago: University of Chicago.

Carroll, W. (1998). Geometric knowledge of middle school students in a reform-based mathematics curriculum. *School Science and Mathematics, 98,* 188–197.

Carroll, W. M. (1993). *Mathematical knowledge of kindergarten and first-grade students in Everyday Mathematics* [UCSMP report]. Unpublished manuscript.

Carroll, W. M. (1994-1995). *Third grade Everyday Mathematics students' performance on the 1993 and 1994 Illinois state mathematics test*. Unpublished manuscript.

Carroll, W. M. (1996a). *A follow-up to the fifth-grade field test of Everyday Mathematics: Geometry, and mental and written computation*. Chicago: University of Chicago School Mathematics Project.

Carroll, W. M. (1996b). Use of invented algorithms by second graders in a reform mathematics curriculum. *Journal of Mathematical Behavior, 15,* 137–150.

Carroll, W. M. (1996c). Mental computation of students in a reform-based mathematics curriculum. *School Science and Mathematics, 96,* 305–311.

Carroll, W. M. (1997). Mental and written computation: Abilities of students in a reform-based mathematics curriculum. *Mathematics Educator, 2*(1), 18–32.

Carroll, W. M. (2000). Invented computational procedures of students in a standards-based curriculum. *Journal of Mathematical Behavior, 18,* 111–121.

Carroll, W. M. (2001a). *A longitudinal study of children in the Everyday Mathematics curriculum*. Unpublished manuscript.

Carroll, W. M. (2001b). Students in a standards-based curriculum: Performance on the 1999 Illinois State Achievement Test. *Illinois Mathematics Teacher, 52*(1), 3–7.

Carroll, W. M., & Fuson, K. C. (1998). *Multidigit computation skills of second and third graders in Everyday Mathematics: A follow-up to the longitudinal study*. Unpublished manuscript.

Carroll, W. M., & Isaacs, A. (2003). Achievement of students using the University of Chicago School Mathematics Project's Everyday Mathematics. In S. L. Senk & D. R. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do students learn?* (pp. 79–108). Mahwah, NJ: Lawrence Erlbaum.

Carroll, W. M., & Porter, D. (1994). *A field test of fourth grade Everyday Mathematics: Summary report.* Chicago: University of Chicago School Mathematics Project, Elementary Component.

Carter, A., Beissinger, J., Cirulis, A., Gartzman, M., Kelso, C., & Wagreich, P. (2003). Student learning and achievement with math. In S. L. Senk & D. R. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do students learn?* (pp. 45–78). Mahwah, NJ: Lawrence Erlbaum.

Chambers, E. A. (2003). *Efficacy of educational technology in elementary and secondary classrooms: A meta-analysis of the research literature from 1992–2002*. Unpublished doctoral dissertation, Southern Illinois University at Carbondale.

Chan, L., Cole, P., & Cahill, R. (1988). Effect of cognitive entry behavior, mastery level, and information about criterion on third graders' mastery of number concepts. *Journal for Research in Mathematics Education, 19,* 439–448.

Chang, K. E., Sung, Y. T., & Lin, S. F. (2006). Computer-assisted learning for mathematical problem solving. *Computers and Education, 46*(2), 140–151.

Chase, A., Johnston, K., Delameter, B., Moore, D., & Golding, J. (2000). *Evaluation of the Math Steps curriculum: Final report.* Boston: Houghton Mifflin.

Cheung, A., & Slavin, R. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Education Research, 75,* 247–284.

Chiang, A. (1978) *Demonstration of the use of computer-assisted instruction with handicapped children: Final report.* (Rep. No. 446-AM-66076A). Arlington, VA: RMC Research Corp. (ERIC Document Reproduction Service No. ED116913)

Clariana, R. B. (1994). *The effects of an integrated learning system on third graders' mathematics and reading achievement.* San Diego, CA: Jostens Learning Corporation. (ERIC Document Reproduction Service No. ED409181)

Clariana, R. B. (1996). Differential achievement gains for mathematics computation, concepts, and applications with an integrated learning system. *Journal of Computers in Mathematics and Science, 15*(3), 203–215.

Clark, R. E. (Ed.). (2001). *Learning from media: Arguments, analysis, and evidence.* Greenwich, CT: Information Age.

Clarke, B., & Shinn, M. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33,* 234–248.

Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., et al. (1991). Assessment of a problem-centered second-grade mathematics project. *Journal for Research in Mathematics Education, 22*(1), 3–29.

Cognition and Technology Group at Vanderbilt. (1992). The Jasper series as an example of anchored instruction: Theory, program description, and assessment data. *Educational Psychologist, 27*(3), 291–315.

Comprehensive School Reform Quality Center. (2005). *CSRQ Center Report on Elementary School Comprehensive School Reform Models.* Washington, DC: American Institutes for Research.

Confrey, J. (2006). Comparing and contrasting the National Research Council report on evaluating curricular effectiveness with the What Works Clearinghouse approach. *Educational Evaluation and Policy Analysis, 28*(3), 195–213.

Cooper, H. (1998). *Synthesizing research* (3rd ed.). Thousand Oaks, CA: Sage.

Cooperative Mathematics Project. (1996). *Findings from the evaluation of the Number Power Mathematics Program: Final report to the National Science Foundation*. Oakland, CA: Developmental Studies Center.

Cox, B. F. (1986). *Instructional management system: How it affects achievement, self-concept, and attitudes toward math of fifth-grade students*. Unpublished doctoral dissertation, Oklahoma State University.

Craig, J., & Cairo, L. (2005). *Assessing the relationship between questioning and understanding to improve learning and thinking (QUILT) and student achievement in mathematics: A pilot study*. Charleston, WV: Appalachia Educational Lab.

Cramer, K. A., Post, T. R., & delMas, R. C. (2002). Initial fraction learning by fourth and fifth grade students: A comparison of the effects of using commercial curricula with the effects of using the Rational Number Project curriculum. *Journal for Research in Mathematics Education, 33*(2), 111–144.

Crawford, D. B., & Snider, V. E. (2000). Effective mathematics instruction: The importance of curriculum. *Education and Treatment of Children, 23*(2), 122–142.

Crenshaw, H. M. (1982). *A comparison of two computer assisted instruction management systems in remedial math*. Unpublished doctoral dissertation, Texas A&M University.

Dahn, V. L. (1992). *The effect of integrated learning systems on mathematics and reading achievement and student attitudes in selected Salt Lake City, Utah, elementary schools*. Unpublished doctoral dissertation, Brigham Young University.

De Russe, J. T. (1999). *The effect of staff development training and the use of cooperative learning methods on the academic achievement of third through sixth grade Mexican-American students*. Unpublished doctoral dissertation, Texas A&M University.

Dev, P. C., Doyle, B. A., & Valenta, B. (2002). Labels needn't stick: "At-risk" first graders rescued with appropriate intervention. *Journal of Education for Students Placed at Risk, 7*(3), 327–332.

Dilworth, R., & Warren, L. (1980). *An independent investigation of Real Math: The field-testing and learner-verification studies.* San Diego, CA: Center for the Improvement of Mathematics Education.

Dobbins, E. R. (1993). *Math computer assisted instruction with remedial students and students with mild learning/behavior disabilities*. Unpublished doctoral dissertation, University of Alabama.

Donnelly, L. (2004, April). *Year two results: Evaluation of the implementation and effectiveness of SuccessMaker during 2002–2003.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, California.

Drueck, J. V., Fuson, K. C., Carroll, W. M., & Bell, M. S. (1995, April). *Performance of U.S. first graders in a reform math curriculum compared to Japanese, Chinese, and traditionally taught U.S. students*. Paper presented at the annual meeting of the American Education Research Association, San Francisco, CA.

DuPaul, G. J., Ervin, R. A., Hook, C. L., & McGoey, K. E. (1998). Peer tutoring for children with attention deficit hyperactivity disorder: Effects on classroom behavior and academic performance. *Journal of Applied Behavior Analysis, 31,* 595–592.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., et al. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort.* Washington, DC: Institute of Education Sciences.

Earnheart, M. (1989). *A study of the effectiveness of mastery learning in a low socioeconomic setting.* Unpublished doctoral dissertation, University of Mississippi.

Ebmeier, H., & Good, T. (1979). The effects of instructing teachers about good teaching on the mathematics achievement of fourth grade students. *American Educational Research Journal, 16*(1), 1–16.

EDSTAR. (2004). *Large-scale evaluation of student achievement in districts using Houghton Mifflin Mathematics: Phase two*. Raleigh, NC: Author.

Ellington, A. J. (2003). A meta-analysis of the effects of calculators on students' achievement and attitude levels in precollege mathematics classes. *Journal for Research in Mathematics Education, 34*(5), 433–463.

Emihovich, C., & Miller, G. (1988). Effects of Logo and CAI on Black first graders' achievement, reflectivity, and self-esteem. *Elementary School Journal, 88*(5), 472–487.

Estep, S. G., McInerney, W. D., Vockell, E., & Kosmoski, G. (1999-2000). An investigation of the relationship between integrated learning systems and academic achievement. *Journal of Educational Technology Systems, 28*(1), 5–19.

Fahsl, A. J. (2001). *An investigation of the effects of exposure to Saxon math textbooks, socioeconomic status and gender on math achievement scores*. Unpublished doctoral dissertation, Oklahoma State University.

Fantuzzo, J. W., Davis, G. Y., & Ginsburg, M. D. (1995). Effects of parent involvement in isolation or in combination with peer tutoring on student self-concept and mathematics achievement. *Journal of Educational Psychology, 87*(2), 272–281.

Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment: A component analysis. *Journal of Educational Psychology, 84*(3), 331–339.

Fantuzzo, J. W., Polite, K., & Grayson, N. (1990). An evaluation of reciprocal peer tutoring across elementary school settings. *Journal of School Psychology, 28,* 309–323.

Faykus, S. P. (1993). *Impact of an integrated learning system: Analysis of student achievement through traditional and curriculum-based procedures*. Unpublished doctoral dissertation, Arizona State University.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*(4), 403–434.

Fischer, F. (1990). A part-part-whole curriculum for teaching number to kindergarten. *Journal for Research in Mathematics Education, 21,* 207–215.

Fletcher, J. D., Hawley, D. E., & Piele, P. K. (1990). Costs, effects, and utility of microcomputer-assisted instruction. *American Educational Research Journal, 27*(4), 783–806.

Florida TaxWatch. (2005). *Florida TaxWatch's comparative evaluation of Project CHILD: Phase IV, 2005*. Retrieved from http:www.floridataxwatch.org/projchild/projchild4.html

Flowers, J. (1998). *A study of proportional reasoning as it relates to the development of multiplication concepts*. Unpublished doctoral dissertation, University of Michigan.

Foley, M. M. (1994). *A comparison of computer-assisted instruction with teacher-managed instructional practices.* Unpublished doctoral dissertation, Columbia University Teachers College.

Follmer, R. (2001). *Reading, mathematics, and problem solving: The effects of direct instruction in the development of fourth grade students' strategic reading and problem-solving approaches to text-based, non-routine mathematics problems*. Unpublished doctoral dissertation, Widener University.

Freiberg, H. J., Connell, M. L., & Lorentz, J. (2001). Effects of consistency management on student mathematics achievement in seven Chapter 1 elementary schools. *Journal of Education for Students Placed at Risk, 6*(3), 249–270.

Freiberg, H. J., Huzinec, C., & Borders, K. (2006). *Classroom management and mathematics student achievement: A study of fourteen inner-city elementary schools*. Unpublished manuscript.

Freiberg, H. J., Prokosch, N., Treiser, E. S., & Stein, T. (1990). Turning around five at-risk elementary schools. *Journal of School Effectiveness and School Improvement, 1*(1), 5–25.

Freiberg, H. J., Stein, T. A., & Huang, S. (1995). Effects of classroom management intervention on student achievement in inner-city elementary schools. *Educational Research and Evaluation, 1*(1), 36–66.

Fuchs, L., Compton, D., Fuchs, D., Paulsen, K., Bryant, J., & Hamlett, C. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*(3), 493–513.

Fuchs, L., Fuchs, D., Finelli, R., Courey, S., & Hamlett, C. (2004). Expanding schema-based transfer instruction to help third graders solve real-life mathematical problems. *American Educational Research Journal, 41*(2), 419–445.

Fuchs, L., Fuchs, D., Finelli, R., Courey, S., Hamlett, C., Sones, E., et al. (2006). Teaching third graders about real-life mathematical problem solving: A randomized control study. *Elementary School Journal, 106*(4), 293–311.

Fuchs, L., Fuchs, D., & Hamlett, C. (1989). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children, 55*(5), 429–438.

Fuchs, L., Fuchs, D., Hamlett, C., & Stecker, P. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28*(3), 617–641.

Fuchs, L., Fuchs, D. & Karns, K. (2001). Enhancing kindergarteners' mathematical development: Effects of peer-assisted learning strategies. *Elementary School Journal, 101*(5), 495–510.

Fuchs, L., Fuchs, D., Karns, K., Hamlett, C., & Katzaroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. *American Educational Research Journal, 36*(3), 609–646.

Fuchs, L., Fuchs, D., Karns, K., Hamlett, C., Katzaroff, M., & Dutka, S. (1997). Effects of task-focused goals on low-achieving students with and without learning disabilities. *American Educational Research Journal, 34*(3), 513–543.

Fuchs, L., Fuchs, D. Phillips, N., Hamlett, C., & Karns, K. (1995). Acquisition and transfer effects of classwide peer-assisted learning strategies in mathematics for students with varying learning histories. *School Psychology Review, 24*(4), 604–620.

Fuchs, L., Fuchs, D., & Prentice, K. (2004). Responsiveness to mathematical problem-solving instruction: Comparing students at risk of mathematics disability with and without risk of reading disability. *Journal of Learning Disabilities, 37*(4), 293–306.

Fuchs, L., Fuchs, D., Prentice, K., Burch, M., Hamlett, C., Owen, R., Hosp, M., et al. (2003). Explicitly teaching for transfer: Effects on third-grade students' mathematical problem solving. *Journal of Educational Psychology, 95*(2), 293–305.

Fuchs, L., Fuchs, D., Prentice, K., Burch, M., Hamlett, C., Owen, R., & Schroeter, K. (2003). Enhancing third-grade students' mathematical problem solving with self-regulated learning strategies. *Journal of Educational Psychology, 95*(2), 306–315.

Fuchs, L., Fuchs, D., Prentice, K., Hamlett, C., Finelli, R., & Courey, S. (2004). Enhancing mathematical problem solving among third-grade students with schema-based instruction. *Journal of Educational Psychology, 96*(4), 635–647.

Fuchs, L., Fuchs, D., Yazdian, L., & Powell, S. (2002). Enhancing first-grade children's mathematical development with peer-assisted learning strategies. *School Psychology Review, 31*(4), 569–583.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Appleton, A. C. (2002). Explicitly teaching for transfer: Effects on the mathematical problem-solving performance of students with mathematical disabilities. *Learning Disabilities Research and Practice, 17*(2), 90–106.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children, 60*(6), 518–537.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., Karns, K., & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *Elementary School Journal, 97*(3), 223–249.

Fueyo, V., & Bushell, D. (1998). Using number line procedures and peer tutoring to improve the mathematics computation of low-performing first graders. *Journal of Applied Behavior Analysis, 31*(3), 417–430.

Fuson, K., Carroll, W., & Drueck, J. (2000). Achievement results for second and third graders using the standards-based curriculum Everyday Mathematics. *Journal for Research in Mathematics Education, 31*(3), 277–295.

Fuson, K. C., & Carroll, W. M. (n.d.-a). *Performance of U.S. fifth graders in a reform math curriculum compared to Japanese, Chinese, and traditionally taught U.S. students*. Unpublished manuscript.

Fuson, K. C., & Carroll, W. M. (n.d.-b). *Summary of comparison of Everyday Math (EM) and McMillan (MC): Evanston student performance on whole-class tests in Grades 1, 2, 3, and 4*. Unpublished manuscript.

Gabbert, B., Johnson, D. W., & Johnson, R. T. (1986). Cooperative learning, group-to-individual transfer, process gain, and the acquisition of cognitive reasoning strategies. *Journal of Psychology, 120*(3), 265–278.

Gallagher, P. A. (1991). *Training the parents of Chapter 1 math students in the use of mastery learning strategies to help increase math achievement*. Unpublished doctoral dissertation, University of Pittsburgh.

Gatti, G. (2004a). ARC study. In Pearson Education (Ed.). *Investigations in Number, Data, and Space: Validation and research.* Upper Saddle River, NJ: Pearson.

Gatti, G. (2004b). *Scott Foresman–Addison Wesley Math national effect size study.* (Available from Pearson Education, K–12 School Group, 1 Lake Street, Upper Saddle River, NJ 07458)

Giancola, S. (2000). *Evaluation results of the Delaware Challenge Grant Project Lead Education Agency: Capital School District*. Newark: University of Delaware.

Gilbert-Macmillan, K. (1983). *Mathematical problem solving in cooperative small groups and whole class instruction.* Unpublished doctoral dissertation, Stanford University.

Gill, B. (1995). *Project CHILD middle school follow-up evaluation: Final report*. Daniel Memorial Institute, Jacksonville, FL.

Gilman, D., & Brantley, T. (1988). *The effects of computer-assisted instruction on achievement, problem-solving skills, computer skills, and attitude: A study of an experimental program at Marrs Elementary School, Mount Vernon, Indiana.* Indianapolis: Indiana State University. (ERIC Document Reproduction Service No. ED302232)

Ginsburg, A., Leinwand, S., Anstrom, T., & Pollok, E. (2005). *What the United States can learn from Singapore's world-class mathematics system (and what Singapore can learn from the United States): An exploratory study*. Washington, DC: American Institutes for Research.

Ginsburg-Block, M. (1998). *The effects of reciprocal peer problem-solving on the mathematics achievement, academic motivation and self-concept of "at risk" urban elementary students*. Unpublished doctoral dissertation, University of Pennsylvania.

Ginsburg-Block, M., & Fantuzzo, J. W. (1997). Reciprocal peer tutoring: An analysis of "teacher" and "student" interactions as a function of training and experience. *School Psychology Quarterly, 12,* 134–149.

Ginsburg-Block, M. D., & Fantuzzo, J. W. (1998). An evaluation of the relative effectiveness of NCTM standards-based interventions for low-achieving urban elementary students. *Journal of Educational Psychology, 90,* 560–569.

Glassman, P. (1989). *A study of cooperative learning in mathematics, writing, and reading in the intermediate grades: A focus upon achievement , attitudes, and self-esteem by gender, race, and ability group*. Unpublished doctoral dissertation, Hofstra University.

Glazerman, S., Levy, D. M., & Myers, D. (2002). *Nonexperimental replications of social experiments: A systematic review.* Washington, DC: Mathematica.

Goldberg, L. F. (1989). *Implementing cooperative learning within six elementary school learning disability classrooms to improve math achievement and social skills.* Unpublished doctoral dissertation, Nova University.

Good, K., Bickel, R., & Howley, C. (2006). *Saxon Elementary Math program effectiveness study*. Charleston, WV: Edvantia.

Good, T. L. & Grouws, D. A. (1979). The Missouri Mathematics Effectiveness Project: An experimental study in fourth-grade classrooms. *Journal of Educational Psychology, 71*(3), 355–362.

Goodrow, A. (1998). *Children's construction of number sense in traditional, constructivist, and mixed classrooms*. Unpublished doctoral dissertation, Tufts University.

Great Source. (2006). *Every Day Counts Grades K–6: Research base and program effectiveness* [computer manual]. Retrieved from http://www.greatsource.com/GreatSource/pdf/EveryDayCountsResearch206.pdf

Greenwood, C. R., Delquadri, J. C., & Hall, R. V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology, 81*(3), 371–383.

Greenwood, C. R., Dinwiddie, G., Terry, B., Wade, L., Stanley, S., Thibadeau, S., et al. (1984). Teacher-versus-peer mediated instruction: An ecobehavioral analysis of achievement outcomes. *Journal of Applied Behavior Analysis, 17,* 521–538.

Greenwood, C. R., Terry, B., Utley, C. A., & Montagna, D. (1993). Achievement, placement, and services: Middle school benefits of ClassWide Peer Tutoring used at the elementary school. *School Psychology Review, 22*(3), 497–516.

Griffin, S. A., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 25–49). Cambridge, MA: MIT Press.

Grossen, B., & Ewing, S. (1994). Raising mathematics problem-solving performance: Do the NCTM teaching standards help? Final report. *Effective School Practices, 13*(2), 79–91.

Gwaltney, L. (2000). *Year three final report the Lightspan Partnership, Inc. Achieve Now Project: Unified School District 259, Wichita Public Schools*. Wichita, KS: Allied Educational Research and Development Services.

Hallmark, B. W. (1994). *The effects of group composition on elementary students' achievement , self-concept, and interpersonal perceptions in cooperative learning groups*. Unpublished doctoral dissertation, University of Connecticut.

Hansen, E., & Greene, K. (n.d.). *A recipe for math. What's cooking in the classroom: Saxon or traditional?* Retrieved March 1, 2006, from http://www.secondaryenglish.com/recipeformath.html

Hativa, N. (1998). Computer-based drill and practice in arithmetic: Widening the gap between high- and low-achieving students. *American Educational Research Journal, 25*(3), 366–397.

Haynie, T. R. (1989). *The effects of computer-assisted instruction on the mathematics achievement of selected groups of elementary school students*. Unpublished doctoral dissertation, George Washington University.

Heller, L. R., & Fantuzzo, J. W. (1993). Reciprocal peer tutoring and parent partnership: Does parent involvement make a difference? *School Psychology Review, 22*(3), 517–535.

Hess, R. D., & McGarvey, L. J. (1987). School-relevant effects of educational use of microcomputers in kindergarten classrooms and homes. *Journal of Educational Computing Research, 3*(3), 269–287.

Hickey, D. T., Moore, A. L., & Pellegrino, J. W. (2001). The motivational and academic consequences of elementary mathematics environments: Do constructivist innovations and reforms make a difference? *American Educational Research Journal, 38*(3), 611–652.

Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal, 30*(2), 393–425.

Hill, H. (2004). Professional development standards and practices in elementary school math. *Elementary School Journal, 104*(3), 215–232.

Hohn, R., & Frey, B. (2002). Heuristic training and performance in elementary mathematical problem solving. *Journal of Educational Research, 95*(6), 374–380.

Holmes, C. T., & Brown, C. L. (2003). *A controlled evaluation of a total school improvement process, School Renaissance*. Paper presented at the National Renaissance Conference, Nashville, TN.

Hooper, S. (1992). Effects of peer interaction during computer-based mathematics instruction. *Journal of Education Research, 85*(3), 180–189.

Hotard, S., & Cortez, M. (1983). *Computer-assisted instruction as an enhancer of remediation*. Lafayette: University of Southwestern Louisiana.

Houghton Mifflin. (n.d.). *Knowing Mathematics field test results for Lincoln Public Schools: Fall 2002–Spring 2003*. Boston: Author.

Hunter, C. T. L. (1994). *A study of the effect of instructional method on the reading and mathematics achievement of Chapter 1 students in rural Georgia*. Unpublished doctoral dissertation, South Carolina State University.

Interactive. (2000). *Documenting the effects of Lightspan Achieve Now! in the Adams County School District 50: Year 3 report*. Huntington, NY: Lightspan.

Interactive. (2001). *Documenting the effects of Lightspan Achieve Now! in the Hempstead Union Free School District: Year 2 report*. Huntington, NY: Lightspan.

Interactive. (2003, August). *An analysis of CompassLearning student achievement outcomes in Pocatello, Idaho, 2002–03*. (Available from CompassLearning, 9920 Pacific Heights Blvd., San Diego, CA 92121)

Interactive & Metis Associates. (2002). *A multi-year analysis of the outcomes of Lightspan Achievement Now in the Cleveland Municipal School District*. Huntington, NY: Lightspan.

Isbell, S. K. (1993). *Impact on learning of computer-assisted instruction when aligned with classroom curriculum in second-grade mathematics and fourth-grade reading*. Unpublished doctoral dissertation, Baylor University.

Jamison, M. B. (2000). *The effects of linking classroom instruction in mathematics to a prescriptive delivery of lessons on an integrated learning system*. Unpublished doctoral dissertation, Texas Tech University.

Jitendra, A. K., Griffin, C. C., McGoey, K., Gardill, M. C., Bhat, P., & Riley, T. (1998). Effects of mathematical word problem solving by students at risk or with mild disabilities. *Journal of Educational Research, 91*(6), 345–355.

Jitendra, A. K., & Hoff, K. (1996). The effects of schema-based instruction on the mathematical word-problem-solving performance of students with learning disabilities. *Journal of Learning Disabilities, 29*(4), 422–431.

Johnson, D. W., Johnson, R. T., & Scott, L. (1978). The effects of cooperative and individualized instruction on student attitudes and achievement. *Journal of Social Psychology, 104,* 207–216.

Johnson, J., Yanyo, L., & Hall, M. (2002). *Evaluation of student math performance in California school districts using Houghton Mifflin Mathematics*. Raleigh, NC: EDSTAR.

Johnson, L. C. (1985, December). *The effects of the groups of four cooperative learning model on student problem-solving achievement in mathematics*. Unpublished doctoral dissertation, University of Houston.

Johnson-Scott, P. L. (2006). *The impact of Accelerated Math on student achievement.* Unpublished doctoral dissertation, Mississippi State University.

Karper, J., & Melnick, S. A. (1993). The effectiveness of team accelerated instruction on high achievers in mathematics. *Journal of Instructional Psychology, 20*(1), 49–54.

Kastre, N. J. (1995). *A comparison of computer-assisted instruction and traditional practice for mastery of basic multiplication facts.* Unpublished doctoral dissertation, Arizona State University.

Kelso, C. R., Booton, B., & Majumdar, D. (2003). *Washington State Math Trailblazers student achievement report.* Chicago: University of Illinois at Chicago.

Kersh, M. (1970). *A strategy for mastery learning in fifth grade arithmetic.* Unpublished doctoral dissertation, University of Chicago.

King, F. J., & Butzin, S. (1992). An evaluation of Project CHILD. *Florida Technology in Education Quarterly, 4*(4), 45–63.

Kirk, V. C. (2003). *Investigation of the impact of integrated learning system use on mathematics achievement of elementary students*. Unpublished doctoral dissertation, East Tennessee State University.

Kopecky, C. L. (2005). *A case study of the Math Matters professional development program in one elementary school.* Unpublished doctoral dissertation, University of Southern California.

Kosciolek, S. (2003). *Instructional factors related to mathematics achievement: Evaluation of a mathematics intervention*. Unpublished doctoral dissertation, University of Minnesota.

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development, 42*(2), 7–19.

Kroesbergen, E. H., & Van Luit, J. E. H. (2003). Mathematical interventions for children with special education needs. *Remedial and Special Education, 24,* 97–114.

Kromhout, O. M. (1993). *Evaluation report Project CHILD, 1992–1993*. Tallahassee, FL.

Kromhout, O. M., & Butzin, S. M. (1993). Integrating computers into the elementary school curriculum: An evaluation of nine Project CHILD model schools. *Journal of Research of Computing in Education, 26*(1), 55–70.

Kulik, J. A. (2003). *Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say* (SRI Project No. P10446.001). Arlington, VA: SRI International.

Laub, C. (1995). *Computer-integrated learning system and elementary student achievement in mathematics: An evaluation study*. Unpublished Doctoral Dissertation, Temple University.

Laub, C., & Wildasin, R. (1998). *Student achievement in mathematics and the use of computer-based instruction in the Hempfield School District*. Landsville, PA: Hempfield School District.

Leffler, K. (2001). Fifth-grade students in North Carolina show remarkable math gains. *Summary of Independent Math Research.* Madison, WI: Renaissance Learning.

Leiker, V. (1993). *The relationship between an integrated learning system, reading and mathematics achievement, higher order thinking skills and certain demographic variables: A study conducted in two school districts*. Unpublished doctoral dissertation, Baylor University.

Levy, M. H. (1985). *An evaluation of computer assisted instruction upon the achievement of fifth grade students as measured by standardized tests*. Unpublished doctoral dissertation, University of Bridgeport.

Lin, A., Podell, D. M., & Tournaki-Rein, N. (1994). CAI and the development of automaticity in mathematics skills in students with and without mild mental handicaps. *Computers in the Schools, 11,* 43–58.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Long, V. M. (1991). *Effects of mastery learning on mathematics achievement and attitudes*. Unpublished doctoral dissertation, University of Missouri–Columbia.

Lucker, G. W., Rosenfield, D., Sikes, J., & Aronson, E. (1976). Performance in the interdependent classroom: A field study. *American Educational Research Journal, 13,* 115–123.

Lykens, J. S. (2003). *An evaluation of a standards-based mathematics program (Math Trailblazers) in Delaware's Caesar Rodney School District*. Unpublished doctoral dissertation, Wilmington College.

Mac Iver, M., Kemper, E., & Stringfield, S. (2003). *The Baltimore Curriculum Project: Final report of the four-year evaluation study.* Baltimore: Johns Hopkins University, Center for Social Organization of Schools.

Madden, N. A., & Slavin, R. E. (1983). Cooperative learning and social acceptance of mainstreamed academically handicapped students. *Journal of Special Education, 17*(2), 171–182.

Madden, N. A., Slavin, R. E., & Simons, K. (1997). *MathWings: Early indicators of effectiveness* (Rep. No. 17). Baltimore: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.

Madden, N. A., Slavin, R. E., & Simons, K. (1999). *MathWings: Effects on student mathematics performance.* Baltimore: Johns Hopkins University Center for Research on the Education of Students Placed at Risk.

Mahoney, S. J. (1990). *A study of the effects of EXCEL Math on mathematics achievement of second- and fourth-grade students*. Unpublished doctoral dissertation, Northern Arizona University.

Mann, D., Shakeshaft, C., Becker, J., & Kottkamp, R. (1999). *West Virginia story: Achievement gains from a statewide comprehensive instructional technology program.* Beverly Hills, CA: Milken Family Foundation with the West Virginia Department of Education, Charleston.

Manuel, S. Q. (1987). *The relationship between supplemental computer assisted mathematics instruction and student achievement*. Unpublished doctoral dissertation, University of Nebraska.

Martin, J. M. (1986). *The effects of a cooperative, competitive, or combination goal structure on the mathematics performance of Black children from extended family backgrounds.* Unpublished doctoral dissertation, Howard University.

Mason, D. A., & Good, T. L. (1993). Effects of two-group and whole-class teaching on regrouped elementary students' mathematics achievement. *American Educational Research Journal, 30*(2), 328–360.

Math Learning Center. (2003). *Bridges in the classroom: Teacher feedback, student data, & current research*. Salem, OR: Author.

Mathematics Evaluation Committee. (1997). *Chicago mathematics evaluation report: Year two.* Pennington, NJ: Hopewell Valley Regional School District.

Mayo, J. C. (1995). *A quasi-experimental study comparing the achievement of primary grade students using Math Their Way with primary grade students using traditional instructional methods.* Unpublished doctoral dissertation, University of South Carolina.

McCabe, K. J. (2001). *Mathematics in our schools: An effort to improve mathematics literacy*. Unpublished master's thesis, California State University, Long Beach.

McCormick, K. (2005). *Examining the relationship between a standards-based elementary mathematics curriculum and issues of equity*. Unpublished doctoral dissertation, Indiana University.

McDermott, P. A., & Watkins, M. W. (1983). Computerized vs. conventional remedial instruction for learning-disabled pupils. *Journal of Special Education, 17*(1), 81–88.

McKernan, M. M. (1992). *The effects of "Mathematics Their Way" and Chicago Math Project on mathematical applications and story problem strategies of second graders.* Unpublished doctoral dissertation, Drake University.

McWhirt, R., Mentavlos, M., Rose-Baele, J. S., & Donnelly, L. (2003). *Evaluation of the implementation and effectiveness of SuccessMaker.* Charleston, SC: Charleston County School District.

Mehrens, W. A., & Phillips, S. E. (1986). Detecting impacts of curricular differences in achievement test data. *Journal of Educational Measurement, 23*(3), 185–196.

Mercer, C. D., & Miller, S. P. (1992). Teaching students with learning problems in math to acquire, understand, and apply basic math facts. *Remedial and Special Education, 13*(3), 19–35.

Merrell, M. L. B. (1996). *The achievement and self-concept of elementary grade students who have received Direct Instruction*. Unpublished doctoral dissertation, University of Southern Mississippi.

Mevarech, Z. R. (1985). The effects of cooperative mastery learning strategies on mathematics achievement. *Journal of Educational Research, 78*(3), 372–377.

Mevarech, Z. R. (1991). Learning mathematics in different mastery environments. *Journal of Education Research, 84*(4), 225–231.

Mevarech, Z. R., & Rich, Y. (1985). Effects of computer-assisted mathematics instruction on disadvantaged pupils' cognitive and affective development. *Journal of Educational Research, 79*(1), 5–11.

Meyer, L. (1984). Long-term academic effects of Direct Instruction Project Follow Through. *Elementary School Journal, 84*(4), 380–394.

Miller, H. (1997). Quantitative analyses of student outcome measures. *International Journal of Education Research, 27*(2), 119–136.

Mills, S. C. (1997). *Implementing integrated learning systems in elementary classrooms.* Unpublished doctoral dissertation, University of Oklahoma.

Mintz, K. S. (2000). *A comparison of computerized and traditional instruction in elementary mathematics*. Unpublished doctoral dissertation, University of Alabama.

Mokros, J., Berle-Carmen, M., Rubin, A., & O'Neil, K. (1996, April). *Learning operations: Invented strategies that work.* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Mokros, J., Berle-Carmen, M., Rubin, A., & Wright, T. (1994). *Full year pilot Grades 3 and 4: Investigations in numbers, data, and space.* Cambridge, MA: Technical Education Research Centers.

Monger, C. T. (1989). *Effects of a mastery learning strategy on elementary and middle school mathematics students' achievement and subject related affect*. Unpublished doctoral dissertation, University of Tulsa.

Moody, E. C. (1994). *Implementation and integration of a computer-based integrated learning system in an elementary school*. Unpublished doctoral dissertation, Florida State University.

Morgan, J. C. (1994). *Individual accountability in cooperative learning groups: Its impact on achievement and attitude with grade three students*. Unpublished master's thesis, University of Manitoba.

Moskowitz, J. M., Malvin, J. H., Schaeffer, G. A., & Schaps, E. (1983). Evaluation of a cooperative learning strategy. *American Educational Research Journal, 20*(4), 687–696.

Moss, J., & Case, R. (1999). Developing children's understanding of the rational numbers: A new model and an experimental curriculum. *Journal for Research in Mathematics Education, 30*(2), 122–147.

Murphy, L. A. (1998). *Learning and affective issues among higher- and lower-achieving third-graders in math reform classrooms: Perspectives of children, parents, and teachers*. Unpublished doctoral dissertation, Northwestern University.

Murphy, R., Penuel, W., Means, B., Korbak, C., Whaley, A., & Allen, J. (2002). *E-DESK: A review of recent evidence on discrete educational software* (SRI International Report). Menlo Park, CA: SRI International.

National Assessment of Educational Progress. (2005). *The nation's report card.* Washington, DC: National Center for Education Statistics.

National Research Council. (2004). *On evaluating curricular effectiveness.* Washington, DC: National Academies Press.

Nattiv, A. (1994). Helping behaviors and math achievement gain of students using cooperative learning. *Elementary School Journal, 94*(3), 285–297.

New Century Education Corporation. (2003). *New Century Integrated Instructional System, evidence of effectiveness: Documented improvement results from client schools and districts*. Piscataway, NJ: Author.

Nguyen, K. (1994). *The 1993–1994 Saxon mathematics evaluation report.* Oklahoma City: Oklahoma City Public Schools.

Nguyen, K., & Elam, P. (1993). *The 1992–1993 Saxon mathematics evaluation report.* Oklahoma City: Oklahoma City Public Schools.

Opuni, K. A. (2006). *The effectiveness of the Consistency Management & Cooperative Discipline® (CMCD) model as a student empowerment and achievement enhancer: The experiences of two K–12 inner-city school systems*. Paper presented at the 4th Annual Hawaii International Conference of Education, Honolulu, Hawaii.

Orabuchi, I. I. (1992). *Effects of using interactive CAI on primary grade students' higher-order thinking skills: Inferences, generalizations, and math problem solving*. Unpublished doctoral dissertation, Texas Woman's University.

Orr, C. (1991). *Evaluating restructured elementary classes: Project CHILD summative evaluation.* Paper presented at the Southeast Evaluation Association, Tallahassee, FL.

Patterson, D. (2005). *The effects of Classworks in the classroom.* Stephenville, TX: Tarleton State University.

Pellegrino, J. W., Hickey, D., Heath, A., Rewey, K., & Vye, N. J. (1992). *Assessing the outcomes of an innovative instructional program: The 1990–1991 implementation*

of the "Adventures of Jasper Woodbury." Nashville, TN: Vanderbilt University, Learning Technology Center.

Perkins, S. A. (1987). *The relationship between computer-assisted instruction in reading and mathematics achievement and selected student variables*. Unpublished doctoral dissertation, University of Michigan.

Peterson, P. L., Janicki, T. C., & Swing, S. R. (1981). Ability × Treatment interaction effects on children's learning in large-group and small-group approaches. *American Educational Research Journal, 18*(4), 453–473.

Phillips, C. K. (2001). *The effects of an integrated computer program on math and reading improvement in grades three through five*. Unpublished doctoral dissertation, University of Tennessee.

Pigott, H. E., Fantuzzo, J. W., & Clement, P. W. (1986). The effects of reciprocal peer tutoring and group contingencies on the academic performance of elementary school children. *Journal of Applied Behavior Analysis, 19,* 93–98.

Podell, D. M., Tournaki-Rein, N., & Lin, A. (1992). Automatization of mathematics skills via computer-assisted instruction among students with mild mental handicaps. *Education and Training in Mental Retardation, 27,* 200–206.

Pratton, J., & Hales, L. W. (1986). The effects of active participation on student learning. *Journal of Educational Research, 79*(4), 210–215.

Quinn, D. W., & Quinn, N. W. (2001a). *Evaluation series: Preliminary study: PLATO elementary math software, Fairview Elementary, Dayton, Ohio*. Bloomington, MN: PLATO Learning.

Quinn, D. W., & Quinn, N. W. (2001b). *Evaluation series: Grades 1–8, Apache Junction Unified School District 43*. Bloomington, MN: PLATO Learning.

Rader, J. B. (1996). *The effect of curriculum alignment between mathematics instruction and CAI mathematics on student performance and attitudes, staff attitudes, and alignment*. Unpublished doctoral dissertation, United States International University.

Ragosta, M. (1983). Computer-assisted instruction and compensatory education: A longitudinal analysis. *Machine-Mediated Learning, 1*(1), 97–127.

Resendez, M., & Azin, M. (2005). *The relationship between using Saxon Elementary and Middle School Math and student performance on Georgia statewide assessments*. Jackson, WY: PRES Associates.

Resendez, M., & Manley, M. A. (2005). *A study on the effectiveness of the 2004 Scott Foresman–Addison Wesley elementary math program*. Jackson, WY: PRES Associates.

Resendez, M.. & Sridharan, S. (2005). *A study on the effectiveness of the 2004 Scott Foresman–Addison Wesley elementary math program: Technical report*. Jackson, WY: PRES Associates.

Resendez, M., & Sridharan, S. (2006). *A study of the effectiveness of the 2005 Scott Foresman–Addison Wesley elementary math program*. Jackson, WY: PRES Associates.

Resendez, M., Sridharan, S., & Azin, M. (2006). *The relationship between using Saxon Elementary School Math and student performance on the Texas statewide assessments*. Jackson, WY: PRES Associates.

Riordan, J., & Noyce, P. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education, 32*(4), 368–398.

Ross, L. G. (2003). *The effects of a standards-based mathematics curriculum on fourth and fifth grade achievement in two Midwest cities*. Unpublished doctoral dissertation, University of Iowa.

Ross, S. M., & Nunnery, J. A. (2005). *The effect of School Renaissance on student achievement in two Mississippi school districts.* Memphis, TN: Center for Research in Educational Policy, University of Memphis.

Roy, J. W. (1993). *An investigation of the efficacy of computer-assisted mathematics, reading, and language arts instruction*. Unpublished doctoral dissertation, Baylor University.

Ruffin, M. R., Taylor, M., & Butts, L. W. (1991). *Report of the 1989–1990 Barrett Math Program* (Report No. 12, Vol. 25). Atlanta, GA: Atlanta Public Schools, Department of Research and Evaluation. (ERIC Document Reproduction Service No. 365508)

Rust, A. L. (1999). *A study of the benefits of math manipulatives versus standard curriculum in the comprehension of mathematical concepts*. Knoxville, TN: Johnson Bible College. (ERIC Document Reproduction Service No. 436395)

Salvo, L. C. (2005). *Effects of an experimental curriculum on third graders' knowledge of multiplication facts*. Unpublished doctoral dissertation, George Mason University.

Schmidt, S. (1991). *Technology for the 21st century: The effects of an integrated distributive computer network system on student achievement*. Unpublished doctoral dissertation, University of La Verne, CA.

Schoenfeld, A. H. (2006). What doesn't work: The challenge and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher, 35*(2), 13–21.

Schreiber, F., Lomas, C., & Mys, D. (1988). *Evaluation of student reading and mathematics: WICAT computer managed instructional program, Salina Elementary School*. Dearborn, MI: Office of Research and Evaluation, Public Schools.

Sconiers, S., Isaacs, A., Higgins, T., McBride, J., & Kelso, C. (2003). *The Arc Center Tri-State Student Achievement Study.* Lexington, MA: COMAP.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin, 105*(2), 309–316.

Shanoski, L. A. (1986). *An analysis of the effectiveness of using microcomputer assisted mathematics instruction with primary and intermediate level students*. Unpublished doctoral dissertation, Indiana University of Pennsylvania.

Shaughnessy, J., & Davis, A. (1998). *Springfield research study: Impact of Opening Eyes and Visual Mathematics curricula*. Portland, OR: Northwest Regional Educational Laboratory.

Shawkey, J. (1989). *A comparison of two first grade mathematics programs: Math Their Way and Explorations*. Unpublished doctoral dissertation, University of the Pacific.

Sheffield, S. (2004). *Evaluation of Houghton Mifflin Math© 2005 in the Chicopee Public School District: Second year results*.

Sheffield, S. (2005). *Evaluation of Houghton Mifflin Math© 2005 in the Chicopee Public School District: First year results*.

Sherwood, R. D. (1991). *The development and preliminary evaluation of anchored instruction environments for developing mathematical and scientific thinking.* Paper presented at the meeting of the National Association for Research in Science Teaching, Lake Geneva, WI. (ERIC Document Reproduction Service No. ED335221)

Shiah, R. L., Mastropieri, M. A., Scruggs, T. E., & Fulk, B. J. (1994-1995). The effects of computer-assisted instruction on the mathematical problem-solving of students with learning disabilities. *Exceptionality, 5*(3), 131–161.

Simpson, N. (2001). *Scott Foresman California Mathematics validation study pretest–posttest results* (Report No. VM-17-3005-CA). (Available from Pearson Scott Foresman, 1415 L Street, Suite 800, Sacramento, CA 95814)

Sinkis, D. M. (1993). *A comparison of Chapter One student achievement with and without computer-assisted instruction*. Unpublished doctoral dissertation, University of Massachusetts.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*(9), 5–11.

Slavin, R. E. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher, 37*(1), 5–14.

Slavin, R. E., & Karweit, N. L. (1985). Effects of whole class, ability grouped, and individualized instruction on mathematics achievement. *American Educational Research Journal, 22*(3), 351–367.

Slavin, R. E., Lake, C., & Groff, C. (2007). *Effective programs for middle and high school mathematics: A best-evidence synthesis*. Baltimore: Johns Hopkins University, Center for Data-Driven Reform in Education. Available at www.bestevidence.org

Slavin, R. E., Leavey, M., & Madden, N. A. (1984). Combining cooperative learning and individualized instruction: Effects on student mathematics achievement, attitudes, and behaviors. *Elementary School Journal, 84*(4), 409–422.

Slavin, R. E., Madden, N. A., & Leavey, M. (1984a). Effects of cooperative learning and individualized instruction on mainstreamed students. *Exceptional Children, 50*(5), 434–443.

Slavin, R. E., Madden, N. A., & Leavey, M. (1984b). Effects of team assisted individualization on the mathematics achievement of academically handicapped and nonhandicapped students. *Journal of Educational Psychology, 76*(5), 813–819.

Sloan, H. A. (1993). *Direct Instruction in fourth and fifth grade classrooms.* Unpublished master's thesis, Purdue University.

Snider, V. E., & Crawford, D. B. (1996). Action research: Implementing Connecting Math Concepts. *Effective School Practices, 15*(2), 17–26.

Snow, M. (1993). *The effects of computer-assisted instruction and focused tutorial services on the academic achievement of marginal learners*. Unpublished doctoral dissertation, University of Miami.

Spencer, T. M. (1999). *The systemic impact of integrated learning systems on mathematics achievement.* Unpublished doctoral dissertation, Eastern Michigan University.

Spicuzza, R., Ysseldyke, J., Lemkuil, A., Kosciolek, S., Boys, C., & Teelucksingh, E. (2001). Effects of curriculum-based monitoring on classroom instruction and math achievement. *Journal of School Psychology, 39*(6), 541–542.

SRA/McGraw-Hill. (2003). *Everyday Mathematics student achievement studies* (Vol. 4). Chicago: Author.

Stallings, J. (1985). A study of implementation of Madeline Hunter's model and its effects on students. *Journal of Educational Research, 78*(6), 325–337.

Stallings, J., & Krasavage, E. M. (1986). Program implementation and student achievement in a four-year Madeline Hunter Follow-Through Project. *Elementary School Journal, 87*(2), 117–138.

Stallings, P., Robbins, P., Presbrey, L., & Scott, J. (1986). Effects of instruction based on the Madeline Hunter model on students' achievement: Findings from a follow-through project. *Elementary School Journal*, *86*(5), 571–587.

Stecker, P., & Fuchs, L. (2000). Expecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research and Practice, 15*(3), 128–134.

Stevens, J. W. (1991). *Impact of an integrated learning system on third-, fourth-, and fifth-grade mathematics achievement*. Unpublished doctoral dissertation, Baylor University.

Stevens, R. J., & Slavin, R. E. (1995). Effects of a cooperative learning approach in reading and writing on handicapped and nonhandicapped students' achievement, attitudes, and metacognition in reading and writing. *Elementary School Journal, 95,* 241–262.
Stone, T. (1996). *The academic impact of classroom computer usage upon middle-class primary grade level elementary school children*. Unpublished doctoral dissertation, Widener University.
Sullivan, F. L. (1989). *The effects of an integrated learning system on selected fourth- and fifth-grade students*. Unpublished doctoral dissertation, Baylor University.
Suppes, P., Fletcher, J. D., Zanotti, M., Lorton, P. V., Jr., & Searle, B. W. (1973). *Evaluation of computer-assisted instruction in elementary mathematics for hearing-impaired students.* Stanford, CA: Institute for Mathematical Studies in the Social Sciences, Stanford University.
Suyanto, W. (1998). *The effects of student teams–achievement divisions on mathematics achievement in Yogyakarta rural primary schools*. Unpublished doctoral dissertation, University of Houston.
Swing, S., & Peterson, P. (1982). The relationship of student ability and small group interaction to student achievement. *American Educational Research Journal, 19,* 259–274.
Tarver, S. G., & Jung, J. A. (1995). A comparison of mathematics achievement and mathematics attitudes of first and second graders instructed with either a discovery-learning mathematics curriculum or a direct instruction curriculum. *Effective School Practices, 14*(1), 49–57.
Taylor, L. (1999). An integrated learning system and its effect on examination performance in mathematics. *Computers and Education, 32*(2), 95–107.
Teelucksingh, E., Ysseldyke, J., Spicuzza, R., & Ginsburg-Block, M. (2001). *Enhancing the learning of English language learners: Consultation and a curriculum based monitoring system.* Minneapolis: University of Minnesota, National Center on Educational Outcomes.
Tieso, C. (2005). The effects of grouping practices and curricular adjustments on achievement. *Journal for the Education of the Gifted, 29*(1), 60–89.
Tingey, B., & Simon, C. (2001). *Relationship study for SuccessMaker levels and SAT-9 in Hueneme 3 Elementary District, school year 2000–2001, with growth analysis* (Pearson Education Technologies). Pt. Hueneme, CA.
Tingey, B. & Thrall, A. (2000). *Duval County Public Schools evaluation report for 1999–2000* (Pearson Education Technologies). Duval County, FL.
Torgerson, C. J. (2006). Publication bias: The Achilles' heel of systematic reviews? *British Journal of Educational Studies, 54*(1), 89–102.
Trautman, T., & Howe, Q. (2004). *Computer-aided instruction in mathematics: Improving performance in an inner city elementary school serving mainly English language learners*. Oklahoma City, OK: American Education Corporation.
Trautman, T. S., & Klemp, R. (2004). *Computer aided instruction and academic achievement: A study of the AnyWhere Learning System in a district wide implementation*. Oklahoma City, OK: American Education Corporation.
Tsuei, M. (2005, June). *Effects of a Web-based curriculum-based measurement system on math achievement: A dynamic classwide growth model*. Paper presented at the National Educational Computing Conference, Philadelphia, PA.
Turner, L. G. (1985). *An evaluation of the effects of paired learning in a mathematics computer-assisted-instruction program.* Unpublished doctoral dissertation, Arizona State University.

Tuscher, L. (1998). *A three-year longitudinal study assessing the impact of the SuccessMaker program on student achievement and student and teacher attitudes in the Methacton School District elementary schools*. Lehigh, PA: Lehigh University.

Underwood, J., Cavendish, S., Dowling, S., Fogelman, K., & Lawson, T. (1996). Are integrated learning systems effective learning support tools? *Computers Education, 26*(1-3), 33–40.

Van Dusen, L. M., & Worthen, B. R. (1994). The impact of integrated learning system implementation on student outcomes: Implications for research and evaluation. *International Journal of Educational Research, 21*(1), 13–24.

Vaughan, W. (2002). Effects of cooperative learning on achievement and attitude among students of color. *Journal of Educational Research, 95*(6), 359–364.

Villasenor, A., & Kepner, H. S. (1993). Arithmetic from a problem-solving perspective: An urban implementation. *Journal for Research in Mathematics Education, 24*(1), 62–69.

Vogel, J. J., Greenwood-Ericksen, A., Cannon-Bowers, J., & Bowers, C.A. (2006). Using virtual reality with and without gaming attributes for academic achievement. *Journal of Research on Technology in Education, 39*(1), 105–118.

Vreeland, M., Vail, J., Bradley, L., Buetow, C., Cipriano, K., Green, C., et al. (1994). Accelerating cognitive growth: The Edison School Math Project. *Effective School Practices, 13*(2), 64–69.

Waite, R. E. (2000). *A study on the effects of Everyday Mathematics on student achievement of third-, fourth-, and fifth-grade students in a large north Texas urban school district*. Unpublished doctoral dissertation, University of North Texas.

Watkins, M. W. (1986). Microcomputer-based math instruction with first-grade students. *Computers in Human Behavior, 2,* 71–75.

Webster, A. H. (1990). *The relationship of computer assisted instruction to mathematics achievement, student cognitive styles, and student and teacher attitudes*. Unpublished doctoral dissertation, Delta State University, Cleveland, MS.

Webster, W. (1995). *The evaluation of Project SEED, 1991–94: Detroit Public Schools*. Detroit, MI: Detroit Public Schools.

Webster, W., & Chadbourn, R. (1996). *Evaluation of Project SEED through 1995.* Dallas, TX: Dallas Public Schools.

Webster, W. J. (1998). *The national evaluation of Project SEED in five school districts: 1997–1998*. Dallas, TX: Webster Consulting.

Webster, W. J., & Chadbourn, R. A. (1989). *Final report: The longitudinal effects of SEED instruction on mathematics achievement and attitudes*. Dallas, TX: Dallas ISD.

Webster, W. J., & Chadbourn, R. A. (1992). *The evaluation of Project SEED: 1990–91.* Dallas, TX: Dallas ISD.

Webster, W. J., & Dryden, M. (1998). *The evaluation of Project SEED: 1997–1998: Detroit Public Schools.* Detroit, MI: Detroit Public Schools.

Weidinger, D. (2006). *The effects of classwide peer tutoring on the acquisition of kindergarten reading and math skills*. Unpublished doctoral dissertation, University of Kansas.

Wellington, J. (1994). Evaluating a mathematics program for adoption: Connecting Math Concepts. *Effective School Practices, 13*(2), 70–75.

Wenglinsky, H. (1998). *Does it compute? The relationship between educational technology and student achievement in mathematics.* Princeton, NJ: Educational Testing Service Policy Information Center.

What Works Clearinghouse. (2005). *Middle school math curricula.* Washington, DC: U.S. Department of Education. Retrieved July 6, 2006, from http://www.whatworks.ed.gov/Topic.asp?tid=03&ReturnPage=default.asp

What Works Clearinghouse. (2006). *Elementary school math.* Washington, DC: U.S. Department of Education. Retrieved September 21, 2006, from http://w-w-c.org/Topic.asp?tid=04&ReturnPage=default.asp

Whitaker, J. C. (2005). *Impact of an integrated learning system on reading and mathematics achievement.* Unpublished doctoral dissertation, Tennessee State University.

White, M. S. (1996). *The effects of TIPS Math activities on fourth grade student achievement in mathematics and on primary caregivers' and teachers' opinions about primary caregivers' involvement in schoolwork*. Unpublished master's thesis, University of Maryland.

Wildasin, R. L. (1994). *A report on ILS implementation and student achievement in mathematics during the 1993–94 school year.* Landisville, PA: Hempfield School District.

Williams, D. (2005). *The impact of cooperative learning in comparison to traditional instruction on the understanding of multiplication in third grade students*. Unpublished doctoral dissertation, Capella University.

Wilson, C. L., & Sindelar, P. T. (1991). Direct Instruction in math word problems: Students with learning disabilities. *Exceptional Children, 57*(6), 512–519.

Wodarz, N. (1994). *The effects of computer usage on elementary students' attitudes, motivation and achievement in mathematics*. Unpublished doctoral dissertation, Northern Arizona University.

Woodward, J., & Baxter, J. (1997). The effects of an innovative approach to mathematics on academically low-achieving students in inclusive settings. *Exceptional Children, 63*(3), 373–389.

Xin, J. F. (1999). Computer-assisted cooperative learning in integrated classrooms for students with and without disabilities. *Information Technology in Childhood Education, 1*(1), 61–78.

Yager, S., Johnson, R. T., Johnson D. W., & Snider, B. (1986). The impact of group processing on achievement in cooperative learning groups. *Journal of Social Psychology, 126*(3), 389–397.

Ysseldyke, J., Betts, J., Thill, T., & Hannigan, E. (2004). Use of an instructional management system to improve mathematics skills for students in Title I programs. *Preventing School Failure, 48*(4), 10–14.

Ysseldyke, J., Spicuzza, R., Kosciolek, S., & Boys, C. (2003). Effects of a learning information system on mathematics achievement and classroom structure. *Journal of Educational Research, 96*(3), 163–173.

Ysseldyke, J., Spicuzza, R., Kosciolek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003). Using a curriculum-based instructional management system to enhance math achievement in urban schools. *Journal of Education for Students Placed at Risk, 8*(20), 247–265.

Ysseldyke, J., & Tardrew, S. (2002). *Differentiating math instruction: A large-scale study of Accelerated Math* [final rep.]. Wisconsin Rapids, WI: Renaissance Learning.

Ysseldyke, J., & Tardrew, S. (2005). *Use of a progress monitoring system to enable teachers to differentiate math instruction*. Minneapolis: University of Minnesota.

Ysseldyke, J., Tardrew, S., Betts, J., Thill, T., & Hannigan, E. (2004). Use of an instructional management system to enhance math instruction of gifted and talented students. *Journal of Education for the Gifted, 27*(4), 293–310.

Ysseldyke, J. E., & Bolt, D. (2006). *Effect of technology-enhanced progress monitoring on math achievement.* Minneapolis: University of Minnesota.

Zollman, A., Oldham, B., & Wyrick, J. (1989). *Effects of computer-assisted instruction on reading and mathematics achievement of Chapter 1 students*. Lexington, KY: Fayette County Public Schools.

Zuber, R. L. (1992). *Cooperative learning by fifth-grade students: The effects of scripted and unscripted techniques*. Unpublished doctoral dissertation, Rutgers University.

Authors

ROBERT E. SLAVIN is director of the Center for Research and Reform in Education at Johns Hopkins University, director of the Institute for Effective Education at the University of York, and cofounder and chairman of the Success for All Foundation; 200 W. Towsontown Boulevard, Baltimore, MD 21204; *rslavin@jhu.edu*.

CYNTHIA LAKE is a research scientist for the Center for Research and Reform in Education at Johns Hopkins University, CDDRE, 200 W. Towsontown Boulevard, Baltimore, MD 21204; *clake5@jhu.edu.* Her research interests include comprehensive school reform, school improvement, and educational research methods and statistics. She is currently a data analyst of meta-analyses of education for the *Best Evidence Encyclopedia*.